UNIVERSITÀ DEGLI STUDI DI MILANO

XXXII Ciclo di Dottorato in Chimica

Dipartimento di Chimica

NEW SYNTHETIC PROCESSES FOR THE INDUSTRIAL PREPARATION OF GENERIC ACTIVE PRINCIPLES

Doctoral thesis for a Dottorato in Apprendistato di Alta Formazione e Ricerca

Luca Senaldi

Matr. R11802

Tutor: Prof.ssa Anna Bernardi

CoTutor: Dott. Daniele Ciceri

Coordinatore: Prof.ssa Emanuela Licandro

Anno Accademico 2018/2019

Table of contents

List of Abbreviations

2-MeTHF = 2-methyltetrahydrofuran

Ac = acetyl

AcOH = acetic acid

ANDA = Abbreviated New Drug Application

API =Active Pharmaceutical Ingredients

CZE = Capillary Zone Electrophoresis

DBA = dibutylamine

DCM = dichloromethane

DMF = *N,N*-dimethylformamide

DMSO = dimethylsulfoxide

DSC = Differential Scanning Calorimetry

DVS = Dynamic Vapor Sorption

EMA = European Medicines Agency

EtOAC = ethyl acetate

EU = Europe

FDA = Food and Drug Administration

FTIR = Fourier Transform Infrared Spectroscopy

GPC = Gel Permeation Chromatography

HPLC = High Performance Liquid Chromatography

ICP-MS = Inductively Coupled - Mass Spectrometry

LC-MS = Liquid Chromatography – Mass Spectroscopy

MALLS = Multi Angle Laser Light Scattering

MGA = 4-O-methylglucuronic acid

Mn = Number Average Molecular Weight

MS = Mass Spectroscopy

Mw = Weight Average Molecular weight

NMP = N-methyl-2-pyrolidone

NMR = Nuclear Magnetic Resonance

PCA = Principal Component Analysis

PD = polydispersity

Py = pyridine

RI = Refractive index

r.t. = room temperature

RT = Retention time

RRT = Relative Retention Time

SP = sulfated polysaccharide

THF = tetrahydrofuran

TLC = Thin Layer Chromatography

US = United States

VEGF = Vascular Endothelial Grow Factor

Xyl = xylose

XRPD = X-ray Powder Diffraction

Introduction

Dottorato in Apprendistato di Alta Formazione e Ricerca

Università degli Studi di Milano and Indena S.p.A. promoted this "Dottorato in Apprendistato di Alta Formazione e Ricerca". The project deals with new synthetic processes for the industrial preparation of generic active principles and involves topics of interest for the company. The research activity was mainly carried out in Indena research center under the joint supervision of Università degli Studi di Milano (Prof. Anna Bernardi) and Indena S.p.A. (Dr. Daniele Ciceri).

Indena S.p.A.

Indena S.p.A. is a company specialized in the preparation of active ingredients for pharmaceuticals, health foods and personal care. The company was founded in 1921 by Carlo Boccaccio Inverni and Biagio Alberto Della Beffa under the name "Inverni Della Beffa S.p.A.". The company grew over time and the botanical derivatives business was reorganized in 1981 under the name "Indena S.p.A". Nowadays, Indena S.p.A. has more than 800 employees, 5 international branches and 4 production sites located in Settala (MI), Palestro (PV), Tours (France), Bangalore (India). The main site in Settala (Figure 1) hosts also the research and development center, a high containment Kilo-Lab and two pilot plants.

Figure 1. Indena production site in Settala.

Indena products are commercialized mainly in Europe (49%) and in USA (28%). The main business sectors are pharmaceutical (65%), health food (30%) and cosmetic (5%).
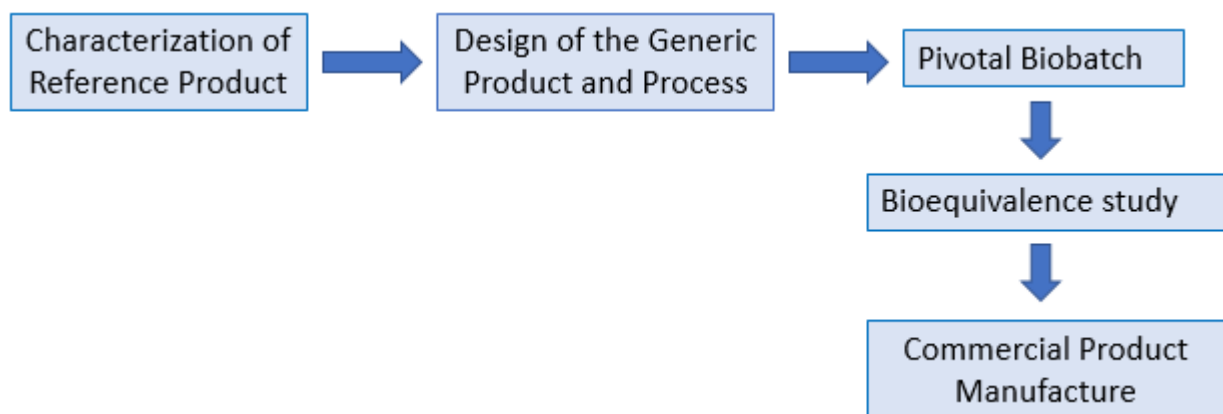
Generic drugs

As mentioned before, this PhD thesis focused on generic drugs and according to definition of the World Health Organization (WHO) a generic drug is a pharmaceutical product, usually intended to be interchangeable with an innovator product, that is manufactured without a license from the innovator company and marketed after the expiry date of the patents or other exclusive rights. Other regulatory agencies, e.g. Food and Drug Administration (FDA) and European Medicines Agency (EMA), can provide slightly different definitions, but the key concept is that a generic drug is the bioequivalent copy of the brand name drug.[1] Two pharmaceutical products can be considered bioequivalent if they are first pharmaceutically equivalent, which means that they have the same active ingredients, dosage form, route of administration and with comparable quality standards.[2] Moreover, the bioavailabilities of two bioequivalent products must be similar to such an extent that their efficacy and safeness are essentially the same.

The introduction of a generic drug on the market can be achieved by abbreviated procedures.[3] For example, in the United States in 1984 Drug Price Competition and Patent Term Restoration Act, commonly known as Hatch-Waxman Act, introduced the so-called ANDA (Abbreviated New Drug Application) for generic medicines. According to this procedure, the generic manufacturer does not have to repeat pre-clinical and clinical studies, but has only to show the bioequivalence to innovator's product. Bioequivalence of two drugs can be claimed when the rate and extent of the absorption of the two drugs are not statistically different (80-125 % of the bioavailability of the originator drug). In Europe the legal situation is more complex because each member has a competent authority in addition to the European Medicines Agency. Anyway, there is a European directive (2001/83/EC), according to which the applicant of a generic medicines does not have to repeat pre-clinical and clinical trials. It is sufficient to demonstrate the bioequivalence of the two products: the generic version must have 80-125% of the bioavailability for common drugs and 90-111 % of the bioavailability for narrow therapeutic index drugs (NTIDs). The different application procedure of a generic medicine is reflected in its development process, which is summarized in Figure 2.[4] The first step is the characterization of the

reference product and this stage can be challenging for complex products like polymeric or supra-molecular structures. Then, the generic product and the process for its preparation are designed and must be not infringing to innovator's patents (e.g. solid form or process patent). The third step is the preparation of a pivotal batch which is used to carry out the bioequivalence study. Once the generic version is demonstrated bioequivalent to innovator product, the generic manufacturer can request the introduction of its version and finally commercialize the drug. The risk of failure in the development of a generic product is usually lower than in the introduction of a new chemical entity because the safety and efficacy of the drug have already been established.

Figure 2. Schematic representation of the development of a generic drug.

The generic drugs are usually between 20% and 90% cheaper than their innovator equivalents. For example, in October 2010 in UK the price of a tablet of generic simvastatin (a cholesterol-lowering medicine) was £0.04 compared to £1.07 of the innovator version.[3] The reduced cost of the generic drugs is strongly correlated to the lower investments necessary for their introduction. Indeed, as mentioned before, the risk of failure is lower and the generic drugs can be introduced through abbreviated procedures which do not require the repetition of the pre-clinical and clinical trials. Moreover, the price of a generic drug can be influenced by local regulations and reimbursement arrangement. For example, in France, the first generic drug price must be 48% lower than originator product and in Italy it must be 20% lower.[5] On the other hand, in the United States and in the United Kingdom there are no specific price regulations and the market is self-regulated by competition. Anyway, in all the countries the generic drugs allow an overall reduction of the cost of the medicines. For

example, it was estimated that in the United States $ 265.1 billion were saved in 2017 thanks to generic drugs (Figure 3).[6]
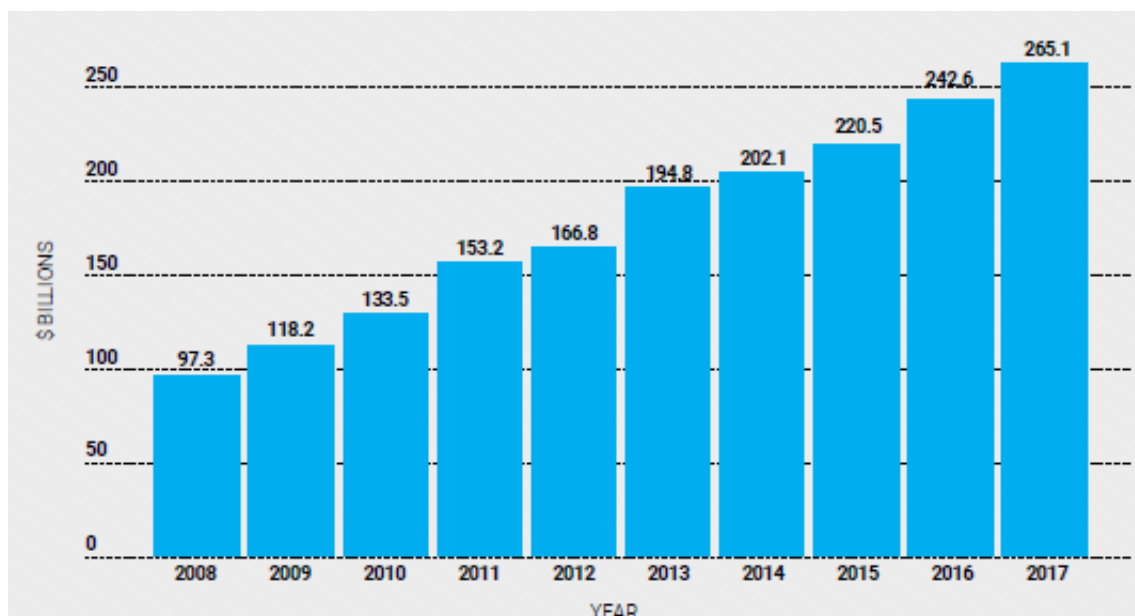
Figure 3. Savings from generics in United States. (Figure modified from Reference 6)

The data show also that there is a constant increase of the savings and the value raised from $97.3 billion in 2008 to $265 billion in 2017. This trend is supported by an increased use of generic drugs as prescription medicines. Indeed, in 1984 in the US 86% of prescription drugs used were brand, while now the major part of them are generics.[7]

Finally, the lower cost of the generic drugs allows to make the medical treatment more accessible and sustainable for the people. It has been shown that it is less frequent the abandon of a therapy with generics (8.1%) than with brand drugs (21.3%).[6]

Aim of the project

As discussed in the previous paragraph, nowadays generic drugs play a key role in the health system of our society and in this PhD project we focused on the design and development of the processes for the preparation of two generic active pharmaceutical ingredients (APIs). In the first part of the project the target API was Lenvatinib mesylate **A** (Figure 4), a potent anti-cancer drug introduced in 2015 by EISAI. The molecule is still protected by several innovator's patents and hence no generic versions are present on the market.
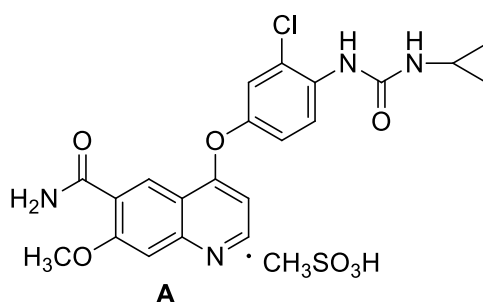
Figure 4. Lenvatinib mesylate.

The development of the generic version of Lenvatinib mesylate will be discussed in section A of this thesis. The project was focused on tackling two crucial issues: the discovery of a new solid form and the development of the preparation process. Indeed, the patent situation of Lenvatinib in the United States revealed the need of a new solid form to anticipate the introduction of the generic version and hence to have a key role on the market. Moreover, the synthetic process had to be developed to satisfy the quality standard required by the regulator agencies for a classical small molecule.

In the second part of PhD (Section B) we focused on a sulfated polysaccharide **B** (Figure 5), a drug with anti-coagulant, fibrinolytic and anti-inflammatory properties.
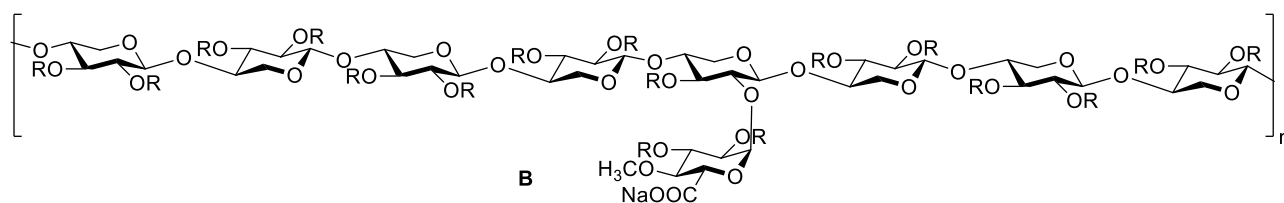
Figure 5. Sulfated polysaccharide. R = SO$_3$Na, H or Ac.

Although the sulfated polysaccharide was introduced several years ago and has no patent protections since 2006, there are no generic versions of this drug on the high regulated markets (United States and Europe). From a commercial point of view the lack of generics and consequently of competition represents a big opportunity. On the other hand, scientifically this is a signal of the difficulties in the reproduction of the same compound. Indeed, the sulfated polysaccharide is a complex API and several important features of its structure are still unknown. Therefore, several orthogonal analytical methods were exploited to investigate and to elucidate the Listed Reference Drug (RLD), the originator sulfated polysaccharide. Contemporarily the process for its preparation was designed from the extraction of a partially acetylated glucuronoxylan from beech wood to the sulfation and product purification. Finally, in the last chapter of Section B a deep comparison of the

7

Originator and Indena version of the sulfated polysaccharide by several analytical techniques will be discussed. In section B a few aspects of the sulfated polysaccharide and its preparation process will not be described for confidentiality reasons.

References

[1] R. Alfonso-Cristancho, T. Andia, T. Barbosa, J. Watanabe, *Appl. Health Econ. Health Policy*, **2015**, 13, S5-S11.

[2] D. Birkett, *Austr. Presc.*, **2003**, 26, 85-87.

[3] S. Dunne, B. Shannon, C. Dunne, W. Cullen, *BMC Pharamacol. Toxicol.*, **2013**, 14, 1-19.

[4] R. Lionberger, *AAPS J.*, **2008**, 10, 103-109.

[5] J. Puig-Junoy, *Pharmacoeconomics*, **2010**, 28, 649-663.

[6] Generic Drug Acces and Savings Report, Association for Accessible Medicines (AAM), **2018**.

[7] Generic drugs report, Center for drug evaluation and Research.

SECTION A: LENVATINIB MESYLATE

1. Introduction

Lenvatinib mesylate **1.Ms** is an anticancer API characterized by a quinoline ring attached to a phenoxy group which bears a cyclopropylurea moiety (Figure 1.1). It was introduced by Eisai, a Japanese pharmaceutical company, and was approved by FDA (Food and Drug administration) for the treatment of differentiated thyroid cancer (DTC), advanced renal cell carcinoma (RCC) and hepatocellular carcinoma (HCC). Goal of our project was the introduction of the generic version of Lenvatinib with a not-infringing polymorph and according to the quality requirements.

1.Ms
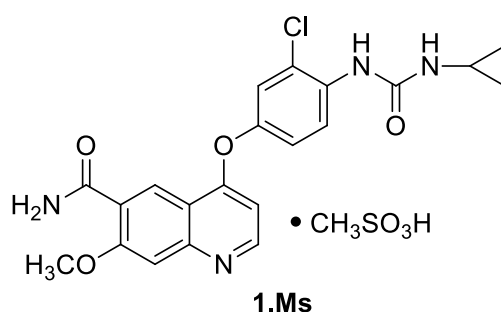
Figure 1.1. Structure of Lenvatinib mesylate.

1.1 Pharmacological properties

The anticancer activity of Lenvatinib mesylate, also known as E7080, is linked to angiogenesis, which plays a crucial role in tumor growth and in the development of metastasis.[1] Angiogenesis is the formation of new blood vessels from pre-existing ones and is regulated by several proangiogenic and antiangiogenic factors. The positive regulators include Vascular Endothelial Grow Factor (VEGF), Fibroblast Grow Factor (FGF) and Platet Derived Grow Factor (PDGF). VEGF is a group of five homodimeric glycoproteins (VEGF-A, VEGF-B, VEGF-C, VEGF-D, PLGF) which interact with tree different tyrosine kinase receptors (VEGFRs).[2] The receptors have different physiological effects: VEGFR-1 is crucial for hematopoietic cell development; VEGFR-2 transmits a signaling cascade to vascular endothelium promoting angiogenesis; VEGFR-3 acts on Lymphatic endothelium and is critical for lymphangiogenesis. Therefore, VEGFR-2 is the principal receptor involved in angiogenesis and its inhibition can down-regulate the formation of new blood vessels blocking the tumor growth and the development of metastasis.
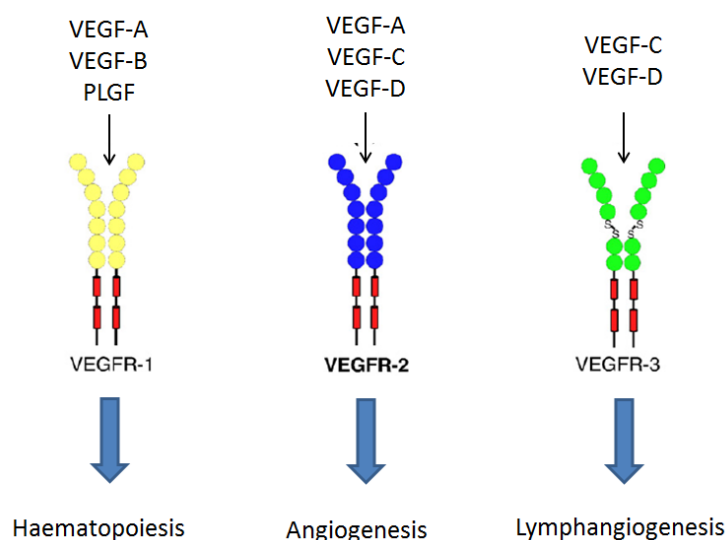
Figure 1.2. Schematic representation of VEGFs interaction with their receptors and the consequent physiological effects. (Figure modified from Reference 2)

Lenvatinib mesylate is an inhibitor of multiple tyrosine kinases, such as the receptors of VEGF, FGF, PDGF and SCF (Stem-cell factor).[3] Among them the highest affinity is towards VEGFR-2 and VEGFR-3 with $IC_{50}s$ of 4.0 and 5.2 nM respectively. Lenvatinib binds VEGFR-2 both at the adenosine-triphosphate (ATP) binding site and at the neighboring allosteric site.[4] Indeed, the quinoline ring and the urea function bind the ATP-binding site, while the cyclopropane ring binds the allosteric region. The interaction of Lenvatinib with VEGFR-2 inhibits the phosphorylation, proliferation and tube formation with a consequent down-regulation of the angiogenesis. Moreover, Lenvatinib is also a good inhibitor of VEGFR-3 and the simultaneous inhibition of VEGFR-2 and VEGFR-3 allows to block the lymphangiogenesis, the formation of new lymphatic vessels, which is essential for the formation of metastases in lymph nodes.[5] Lenvatinib can also target other tyrosine kinase receptors which play important roles in the development of tumors. For example, it has been shown that Lenvatinib inhibits FGFR-1 and PDGFR-β, which are overexpressed in epithelial tumor cells, and consequently inhibits both tumor cell migration and invasion.[6]

All these properties determine the anticancer activity of Lenvatinib mesylate, which was widely studied in phase II and III clinical trials for the treatment of thyroid cancer, renal carcinoma and hepatocellular carcinoma. Thyroid cancer is the fifth most common cancer in USA and the most widespread subgroup is differentiated thyroid carcinoma (DTC), which is usually treated by surgery and radioactive iodine.[7] A phase II clinical study was carried out on the use of Lenvatinib for the treatment of advanced, progressive and iodine-refractory DTC and the data showed an ORR (objective response rate, the portion of patients with a defined regression of tumor) of 50%.[8] The

positive results were confirmed by a double-blind phase III trial, where PFS (progression-free survival, the months lived with the disease without worsening) of Lenvatinib treated group was 18.3 months, compared to 3.6 months of placebo group.[9] Finally, based on these promising results, FDA approved Lenvatinib mesylate for the treatment of differentiated thyroid carcinoma which does not respond to therapy with radioactive iodine.

In 2016 Lenvatinib was approved in combination with everolimus for the treatment of advanced renal cell carcinoma (RCC) based on the results of phase II clinical study. Indeed, according to this trial the combination of Lenvatinib with everolimus increased progression-free survival (PFS) to 14.6 months, compared to 5.5 months of everolimus alone.[10]

Finally, a phase III trial was conducted on patients with unresectable hepatocellular carcinoma (HCC) and Lenvatinb showed a comparable activity to Sorafenib, another anti-cancer drug.[11] Based on these results, in 2018 FDA approved Lenvatinib for first-line treatment of unresectable hepatocellular carcinoma. Similarly, in Europe Lenvatinib was approved by EMA (European Medicines Agency) for the treatment of DTC and HCC and is commercialized under the name Lenvima.

1.2 Patent situation

Lenvatinib mesylate was introduced on the market by EISAI, a Japanese company, which protected the active pharmaceutical ingredient (API) with patents claiming the drug substance, the process, the solid form and the use. When we started the project in 2016, we studied the intellectual property status to understand when the introduction of the generic version would be possible and which patents could be critical. The analysis was carried out for the two main markets. US patent situation will be discussed in paragraph 1.2a, while the EU one in paragraph 1.2b.

1.2a US patent situation

The intellectual property status in the United States is summarized in Figure 1.3 and Lenvatinib mesylate was approved for commercialization on the 13th of February 2015. It has received the new chemical entity (NCE) status, which is proper of active moieties never approved by FDA and entitles an exclusivity period until the 13th of February 2020. Lenvatinib is also considered an orphan drug because it treats a rare medical condition and has an additional exclusivity period which will expire on the 13th of February 2022. Indeed, to promote the development of these not fully profitable drugs the government gives financial incentives and extended exclusivity periods.
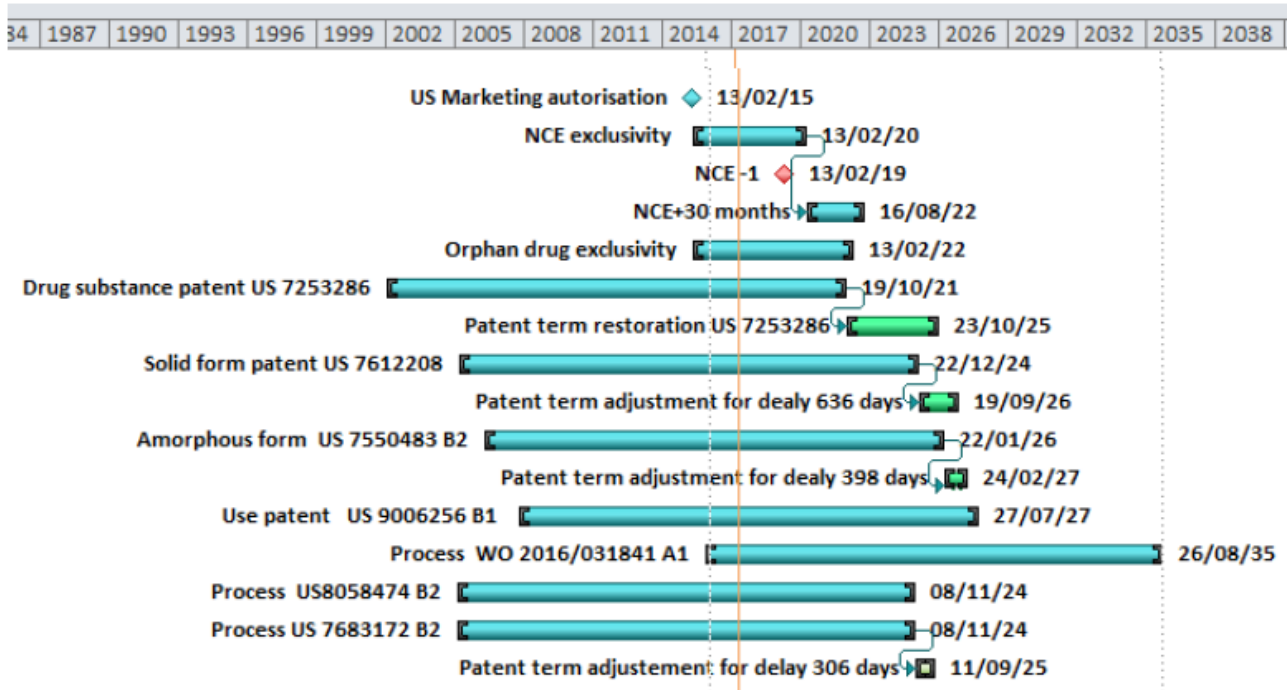
Figure 1.3. US patent situation of Lenvatinib mesylate.

Drug substance is protected by patent US 7253286[12] which will expire on the 23th October 2025, considering the patent term extension (PTE) of 1465 days (about 4 years). PTE was introduced in 1984 by Patent term restoration Act, also known as Hatch – Waxman, which considers the delay between the filing of drug substance patent and the market authorization and hence grants an extension up to five years. The solid form is protected by patent US 7612208[13], while the amorphous form by patent US 7550483[14] and they will expire on the 19th of September 2026 and the 24th of February 2027, respectively. Both the patents have received PTA (Patent term adjustment) extension which is usually given for delays occurred from the filing to patent granting. The drug substance patent will expire before the solid form patent and hence there is a chance for paragraph IV challenge of patent US 76112208 once a new solid form is discovered. Indeed, when you want to introduce a generic drug, you must file an ANDA (Abbreviated New Drug Application) and in a section you must include your position to the Originator drug-related patents. There are four possible scenarios or paragraphs:

- **Paragraph I**: no listed patents
- **Paragraph II**: the drug is already off patents
- **Paragraph III**: the generic manufacturer will stay off until the patents expire
- **Paragraph IV**: one or more patents are invalid or will not be infringed by the manufacturer

Filing an ANDA under Paragraph IV is advantageous because the applicant obtains six months exclusivity as generic manufacturer and consequently it obtains the opportunity to establish a key

role in the generic market. In the case of Lenvatinib a new polymorph is needed to anticipate the introduction of the generic version before the expiration of solid form patent and to apply for a paragraph IV challenge of solid form patent US 7612208. Paragraph IV application must be filed one year before the end of NCE exclusivity, which is the 13th of February 2019.

As far as the preparation process is concerned, there are patent US 8058474[15] and US 7683172[16] which will expire on the 8th of November 2024 and the 11th of September 2025 respectively. As reported in Scheme 1.1, patent US 7683172 claims essentially the synthesis of Lenvatinib by coupling of 4-chloro-quinoline **2** and cyclopropylurea **3**, while patent US 8058474 claims the cyclopropylurea **3** and its phenyl carbamate precursor **4**. The synthesis of 4-chloro-quinoline **2** is reported, but not claimed in patent US 7683172. Anyway, both patents expire before the drug substance patent and hence the claimed process are viable for the preparation of the generic version of Lenvatinib mesylate.
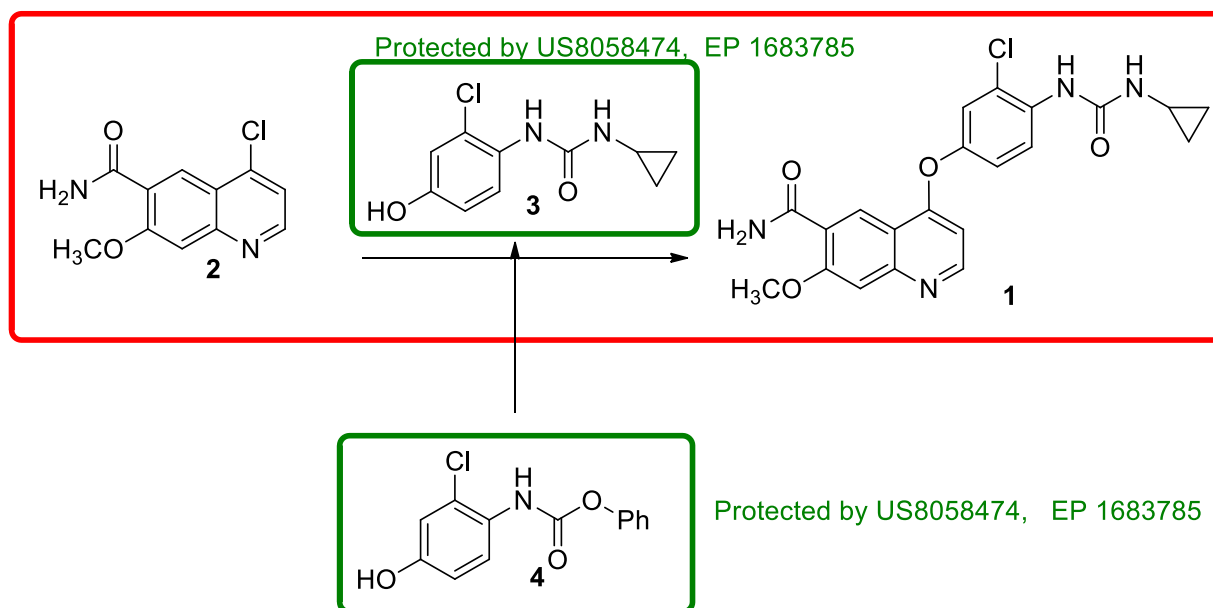
Scheme 1.1. Process claimed by US 805874 and US 7683172.

Finally, patent application WO 2016/031841[17] claims high purity Lenvatinib mesylate and the process for its preparation which is very similar to the one claimed by previous patents.

In conclusion, the generic version can be launched on the US market after the expiration of the drug substance patent on the 23th of October 2025. The preparation process claimed in originator's patents are viable, while a new polymorph is necessary and would allow a paragraph IV challenge.

1.2b EU patent situation

The European patents related to Lenvatinib mesylate are slightly different from US patents and are schematically reported in Figure 1.4.
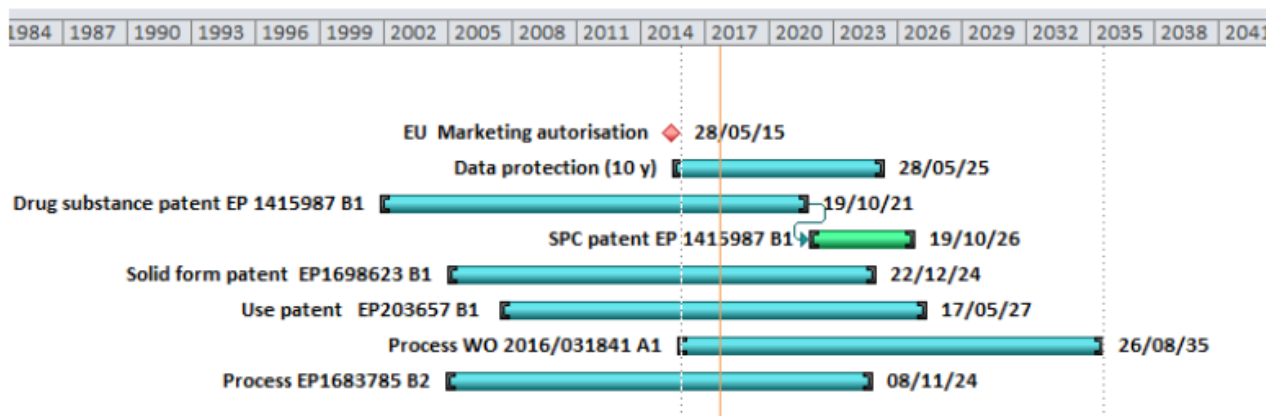

Figure 1.4. EU patent situation of Lenvatinib mesylate.

Lenvatinib mesylate was approved for the commercialization on the 28th of May 2015 and data protection or exclusivity covers the registration data until the 28th of May 2025. In this period an applicant of a generic or biosimilar version can not rely on originator's data to support his application. Drug substance is claimed in patent EP 1415987[12] which will expire on the 19th of October 2026 including SPC extension. SPC (Supplementary Protection Certificate) is an intellectual property right which extends the duration of a patent up to five years, to compensate the long time required to obtain marketing authorization of a drug.

The solid form is covered by patent EP 1698623[13] which will expire on the 22nd of December 2024. Therefore, solid form patent will expire before the drug substance patent and hence in Europe, in contrast to US, there is no need of a new polymorph.

Patent EP 1683785[15,16] protects the synthetic process, which includes both the coupling reaction and the cyclopropylurea intermediate **3** and its precursor **4**, as shown in Scheme 1.2. This patent will expire on the 8th of November 2024, before the expiration of drug substance patent and thus the claimed process is viable for the preparation of our generic version.

It is worth noticing that drug substance patent in Europe will expire about one year after the expiration of the equivalent US patent. Because a patent impedes both the commercialization and the realization of the invention, in that year it will not be possible to produce Lenvatinib mesylate in Europe even if it is commercialized in US. Therefore, it will be necessary to outsource the preparation of API for US market between the 23th of October 2025 and the 19th of October 2026. However, it is possible to prepare registration and validation batches in Europe before drug

substance patent expiration under Bolar provision. Indeed, Bolar provision is originally an US law which is present similarly in Europe and allows the preparation of batches and the collection of data necessary to demonstrate the bioequivalency and to submit the application for a generic drug.

1.3 Aim of the project

Aim of the project is the introduction of the generic version of Lenvatinib mesylate. From US patent situation it is clear that it is necessary a new solid form to anticipate the introduction of the generic drug and to apply for a paragraph IV challenge. This strategy is crucial to obtain a key role in the US generic market of Lenvatinib. Therefore, we worked to discover a new, not infringing polymorph, which will be disclosed in chapter 2. We studied the stability of the new solid form in different temperature and humidity conditions and during mechanical treatment (e.g. micronization). Moreover, we carried out a preliminary evaluation of the solubility of the new polymorph compared to the previous patented ones. Indeed, solubility can be affected by the crystalline structure of a molecule and is the main parameter influencing the bioavailability. A generic drug must be the bioequivalent copy of the branded-name drug and hence the two polymorphs must have comparable solubilities in order to be bioequivalent.

At the same time, we worked on the development of the preparation process of Lenvatinib. Indeed, as described in the previous paragraph, the patented processes are viable both in US and in EU, but they need to be developed in order to be scalable, profitable and especially to satisfy the quality requirements. First of all, our version of Lenvatinib mesylate must have a defined purity and because we do not know the originator's specifications, we must follow the ICH guideline Q3B.[18] According to this guideline, all the known and characterized impurities must be below 0.15%, while all the not identified impurities must be below 0.10%. Therefore, we tried to identify, synthesize and characterize all the most critical impurities to raise the limit to 0.15%. Meanwhile, we developed a process able to minimize the formation of the impurities in the synthetic steps or able to remove them by recrystallization. Furthermore, our API must be in compliance with the ICH guideline Q3C[19], according to which the residual solvents are divided in three different classes and each group of solvents has a defined limit. We tried to design a process able to avoid the use of solvents of class 1 or 2 (solvents to be avoided or limited) or we developed procedures to remove them efficiently.

Finally, the process does not have to match only the quality requirements but also needs to be scalable and profitable. For example, a high number of recrystallizations would allow to reach high purity, but is not sustainable from a productivity and economic point of view.
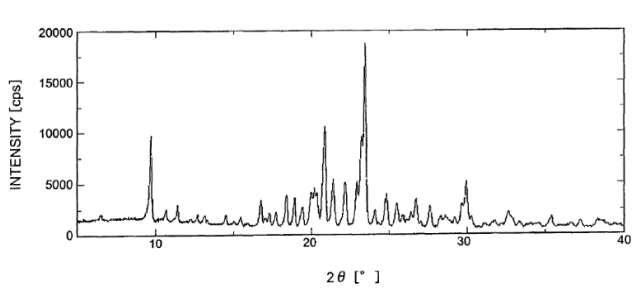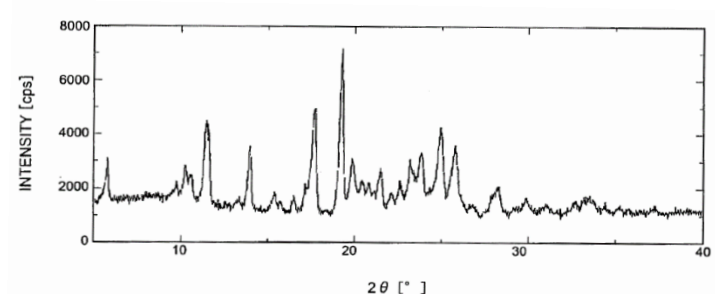
References

[1] D. Hanahan, J. Folkman, *Cell*, **1996**, 86, 353-364.

[2] K. Holmes, O. Roberts, A. Thomas, M. Cross, *Cell. Signal.*, **2007**, 19, 2003-2012.

[3] J. Matsui, Y. Yamamoto, Y. Funahashi, A. Tsuruoka, T. Watanabe, T. Wakabayashi, T. Uenaka, M. Asada, *Int. J. Cancer*, **2008**, 122, 644-671.

[4] Lenvatinib EMA Assessment Report, EMA/250082/2015, 19.

[5] J. Matsui, Y. Funahashi, T. Uenaka, T. Watanabe, A. Tsuruoka, M. Asada, *Clin. Cancer Res.*, **2008**, 14, 5459-5465.

[6] H. Glen, S. Mason, H. Patel, K. Macleod, V. Brunton, *BMC Cancer*, **2011**, 11, 309-318.

[7] M. Cabanillas, D. McFadden, C. Durante, *Lancet*, **2016**, 388, 2783-2795.

[8] M. Cabanillas, M. Schlumberger, B. Jarzab, R. Martins, F. Pacini, B. Robinson, J. McCaffrey, M. Shah, D. Bodenner, D. Topliss, C. Andresen, J. O'Brien, N. Ren, Y. Funahashi, R. Allison, R. Elisei, K. Newbold, L. Licitra, S. Sherman, D. Ball, *Cancer*, **2015**, 121, 2749-2756.

[9] M. Schlumberger, M. Tahara, L. Wirth, B. Robinson, M. Brose, R. Elisei, M. Habra, K. Newbold, M. Shah, A. Hoff, A. Gianoukakis. N. Kiyota, M. Taylor, S. Kim, M. Krzyzanowska, C. Ductus, B. de las Heras, J. Zhu, S. Sherman, *N. Engl. J. Med.*, **2015**, 372, 621-630.

[10] R. Motzer, T. Houtson, H. Glen, M. Michaelson, A. Molina, T. Elsen, J. Jassem, J. Zolnierek, J. Maroto, B. Mellado, B. Melichar, J. Tomasek, A. Kremer, H. Kim, K. Wood, C. Ductus, J. Larkin, *Lancet Oncol.,* **2015**, 16, 1473-1482.

[11] M. Kudo, R. Finn, S. Qin, K. Han, K. Ikeda, F. Piscaglia, A. Baron, J. Park, G. Han, J. Jassem, J. Blanc, A. Vogel, D. Komov, T. Evans, C. Lopez, C. Ductus, M. Guo, K. Saito, S. Kraljevic, T. Tamai, M. Ren, A. Cheng, *Lancet*, **2018**, 391, 1163-1173.

[12] Y. Funahashi, A. Tsuruoka, M. Matsukura, T. Haneda, Y. Fukuda, J. Kamata, K. Takahashi, T. Matsushima, K. Miyazaki, K. Nomoto, T. Watanabe, H. Obaishi, A. Yamaguchi, S. Suzuki, K: Nakamura, F. Mimura, Y. Yamamoto, J. Matsui, K. Matsui, T. Yoshiba, Y. Suzuki, I. Arimoto, **2003**, patent US 7253286, patent EP 1415987.

[13] T. Matsushima, T. Nakamura, K. Yoshizawa, A. Kamada, Y. Ayata, N. Suzuki, I. Arimoto, T. Sakaguchi, M. Gotoda, **2004**, patent US 7612208, patent EP 1698623.

[14] T. Sakaguchi, A. Tsuruoka, **2006**, patent US 7550483.

[15] T. Naito, **2003**, patent US 8058474, patent EP 1683785.

[16] T. Naito, K. Yoshizawa **2002**, patent US 7683172, patent EP 1683785.

[17] T. Nakamura, T. Abe, Y. Miyashita, H. Kuroda, Y. Ayata, A. Akao, **2014**, patent application WO 2016/031841.

[18] ICH harmonized tripartite guideline, Impurities in New drug Products Q3B (R2).

[19] ICH harmonized guideline, Impurities: guideline for Residual Solvents Q3C (R6).

2. Solid Form of Lenvatinib mesylate

As discussed in Chapter 1, a new solid form of Lenvatinib mesylate is necessary to anticipate the introduction of the generic drug and to apply for a paragraph IV challenge in the United States. In this chapter a new solid form will be described and its stability at different temperature and humidity conditions will be discussed. Moreover, we evaluated the effect of mechanical treatment (e.g. micronization) and the solubility compared to known solid forms.

2.1 State of the art

The originator, EISAI, claimed six different solid forms in patent US 7612208 [1]: Form A (anhydrous), Form B (anhydrous), Form C (anhydrous), Form F (hydrate), Form I (acetic acid solvate) and a DMSO (dimethylsulfoxide) solvate. Furthermore, in patent US 7550483 [2] EISAI claimed the amorphous, which is characterized by a flat X-ray diffractogram, and Crystal Pharmatech described another anhydrous crystalline form in patent application WO 2016/184436 [3]. All these solid forms and their XRPD (X-ray Powder diffraction) patterns are summarized in Table 2.1.

Name	Description	XRPD	Patent
Form A	Anhydrous		US 7612208
Form B	Anhydrous		US 7612208
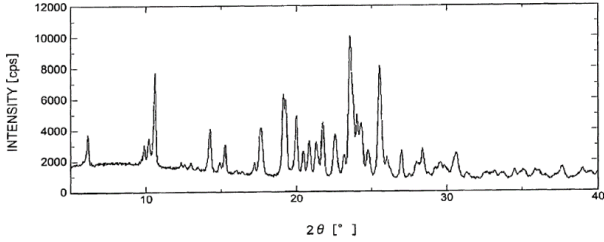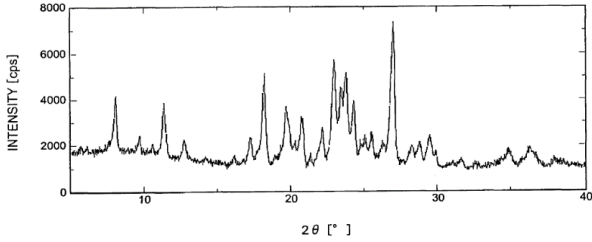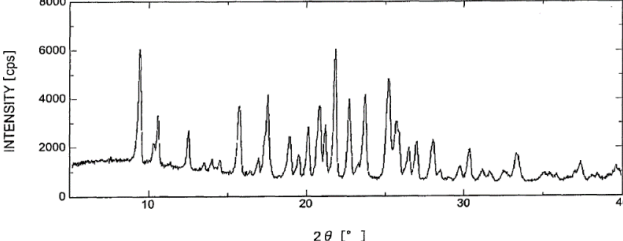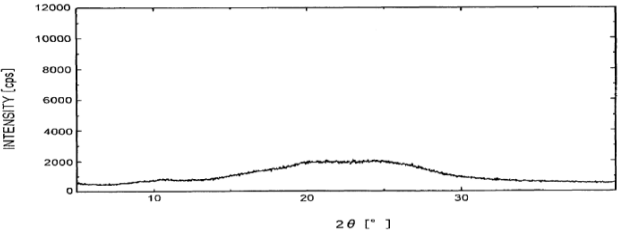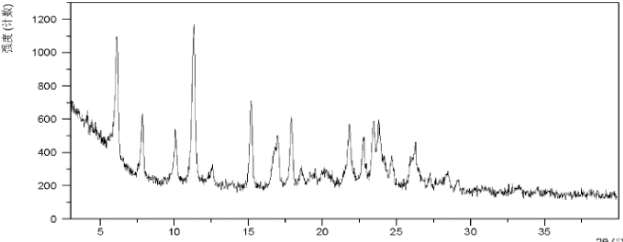
Form C	Anhydrous		US 7612208
Form F	Hydrate		US 7612208
Form I	Acetic acid solvate		US 7612208
No specific name	DMSO solvate	Not reported	US 7612208
Amorphous			US 7550483
No specific name	Anhydrous		WO 2016/184436

Table 2.1. State of art of solid forms of Lenvatinib mesylate and their XRPD patterns.

Nowadays, there are additional crystalline forms claimed in patents WO 2018/196687[4] and WO 2018/122780[5], but they were discovered and patented after the discovery and patenting of our new solid form of Lenvatinib mesylate which will be described in next paragraph.

2.2 Discovery of new solid forms

In order to discover a new crystalline form of Lenvatinib mesylate we collaborated with Polycrystalline, a company specialized in this field, and we started investigating the solubility of the molecule. A visual screening of the solubility was carried out according to the European Pharmacopeia protocol[6] and Lenvatinib showed modest solubility (s < 10 g/L) in several common solvents (e.g. acetonitrile, acetone, chloroform, dichloromethane, ethyl acetate, heptane, methanol, ethanol, tetrahydrofuran, toluene and water). These solvents are not suitable for a monosolvent crystallization, but they can be considered eventually in a mixture as anti-solvent. On the other hand, it has moderate solubility (10 ≤ s < 100 g/L) in *N,N*-dimethylformamide (DMF), *N,N*-dimethylacetamide (DMA), *N*-methyl-2-pyrolidone (NMP), dimethylsulfoxide (DMSO) and acetic acid (AcOH) and has high solubility (s ≥ 100 g/L) in formic acid and trifluoroacetic acid.

Based on the solubility data, several crystallization methods were evaluated and seven new crystal forms were discovered and characterized by XRPD (X-ray Powder Diffraction), which are reported in Figure 2.1, and TGA-EGA (Thermogravimetry - Evolved Gas Analysis).

Figure 2.1. Overlay of XRPD of seven new solid forms of Lenvatinib mesylate.

Dissolution of Lenvatinib mesylate in dimethylsulfoxide and following precipitation by addition of different anti-solvents afforded two new solid forms, named DMSO-1 and DMSO-2. TGA-EGA clarified that DMSO-1 is a not well-defined DMSO and/or water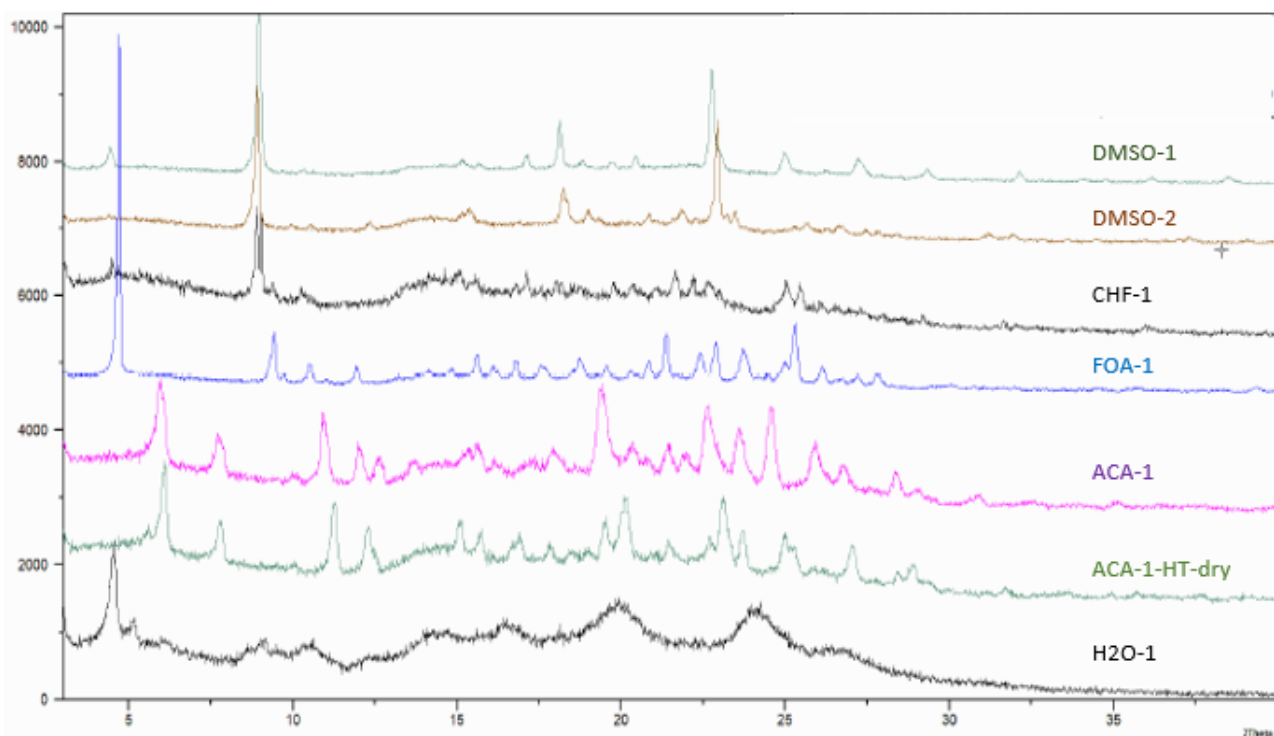 solvate, while DMSO-2 is a DMSO solvate. However, removal of the excess of dimethylsulfoxide is critical for the stability of the polymorphs because washing DMSO-1 crystals with other solvents (e.g. ethyl acetate) determines conversion to patented form I and de-solvation of DMSO-2 by heating above 140°C causes conversion to Form C.

A new form, named CHF-1, was obtained by precipitation of a saturated solution of Lenvatinib mesylate in acetic acid by addition of chloroform. According to TGA-EGA the crystals are solvates of chloroform and/or acetic acid and during de-solvation experiments this new form converts to patented Form B. A similar behavior was observed with another polymorph called FOA-1 which is obtained from a saturated solution of Lenvatinib mesylate in formic acid and addition of different ant-solvents. However, removal of solvents by heating at 130°C causes conversion to originator's Form C.

The evaporation of a saturated acetic acid solution or the addition of an anti-solvent (e.g. ethyl acetate) to a saturated acetic acid solution of Lenvatinib mesylate allows to obtain a new acetic acid solvate, which was called ACA-1. This form converts to a new anhydrous polymorph (ACA-1-HT-dry) by de-solvation at T between 140 and 180°C.

Finally, a low crystalline form, H2O-1, was obtained by evaporation of a suspension of known Form A in water. In conclusion, seven new crystalline forms (DMSO-1, DMSO-2, CHF-1, FOA-1, ACA-1, ACA-1-HT-dry, H2O-1) were obtained and patented in September 2016.[7] Among them, the most promising is ACA-1-HT-dry, which is an anhydrous form and seems to be the most stable. Therefore, we decided to scale-up its preparation from milligrams to grams and to further investigate its properties. In our labs the acetic acid solvate, ACA-1, was obtained by dissolution of Lenvatinib mesylate in 3 V of acetic acid at 60°C followed by addition of 1 V of ethyl acetate and slow decrease of temperature to 40°C. The obtained ACA-1 was dried under vacuum at 60-80°C for 3-5 days to afford the desired polymorph, ACA-1-HT dry. In addition to the above mentioned XRPD and TGA-EGA analysis, the characterization of ACA-1-HT-dry was completed by DSC (Differential Scanning Calorimetry), DVS (Dynamic Vapor Sorption), FTIR (Fourier Transform Infrared Spectroscopy) and Solid State NMR. All the profiles and the analytical details are reported in the experimental section.

DSC is an analytical technique based on heating the sample and comparing its temperature with the temperature of a reference. It is commonly used to characterize a polymorph because the calorimetric phenomena can depend on the crystalline structure of the solid. In the case of Lenvatinib, DSC profile is characterized by a broad endothermic peak below 100°C due to release of moisture and by a melting peak with onset at 179°C, maximum at 191°C and ΔH of about -46 J/g.

DVS (Dynamic Vapor Sorption) is applied to measure the water uptake by modifying the humidity surrounding the sample and measuring the change in weight. The measure is usually carried out increasing the humidity to evaluate water sorption and decreasing the humidity to evaluate the desorption. The difference in water uptake between sorption and desorption is called hysteresis. DVS profiles of Lenvatinib show two main steps: the first one is at 20% of RH (relative humidity) where the sample absorbs about 6% of water which remains almost constant until 90% of RH where water content raises to 13%. According to this profile, ACA-1-HT-dry form of Lenvatinib mesylate can be considered hygroscopic. Moreover, the sorption and desorption of water is not completely reversible and at the end of the first cycle water is about 6% probably due to conversion to hydrate form.

Finally, SS-NMR (Solid State Nuclear Magnetic Resonance) and in particular ^{13}C-CPMAS (Cross Polarization Magic Angle Spinning) NMR of ACA-1-HT-dry form reveals a single set of signals of Lenvatinib structure and a single methyl signal of mesylate group. This means that in the unit cell there are just one independent molecule of Lenvatinib and one of mesylate, indication of a pure crystalline form.

2.3 Stability study of ACA-1-HT-dry form

We evaluated the stability of new solid form ACA-1-HT-dry of Lenvatinib mesylate at different temperature and humidity conditions. We started by exposing a vial of sample at ambient temperature and humidity for 5 days and we did not observe any change in the diffractogram. The same conclusion was reached by placing a vial of sample at 40°C and 75% RH for 20 days. According to these preliminary trials, the new solid form does not convert easily to other polymorphs and hence we decided to study more accurately its stability with accelerated and long-term testing (Table 2.2). For these experiments, Lenvatinib mesylate (ACA-1-HT dry form) was introduced in a closed amber glass bottle under nitrogen atmosphere placed inside a food grade polyethylene bag. This bag was placed inside an Aluminum multilayer thermo-sealed bag, under

vacuum. The packaged samples were placed in three different climatic chambers (5°C, 25°C and 60% RH, 40°C and 75% RH) and were analyzed at different times by XRPD and FTIR for solid state identification and by Karl Fischer titration (KF) to determine the water content.

		Storage conditions and time (months)							
		5±3°C					25±2°C/65±5%R.H.		40±2°C/75±5%R.H.	
	0	3	6	12	18	3	6	3	6
XRPD identification	ACA-1-HT-dry	No change	No change	No change	No change	No change	No change	No change	No change
FTIR identification	ACA-1-HT-dry	No change	No change	No change	No change	No change	No change	No change	No change
Water content (K.F.)	0.45	2.88	1.74	1.50	2.05	2.04	3.70	2.95	6.49

Table 2.2. Stability test of Lenvatinib mesylate - ACA-1-HT-dry form in amber glass under nitrogen + multilayer bag thermo-sealed under vacuum.

According to XRPD and FTIR profiles, there are no changes in the solid form both in long-term test at low temperature and in accelerated conditions at higher temperature and relative humidity (40°C and 75% RH). On the other hand, there are differences in the water content determined by Karl Fischer titration. A slight oscillation of the value is due to the sample preparation before the analysis. Indeed, the sample is weighed on air and according to the time spent in this operation the solid can absorb variable quantities of water. Anyway, especially in accelerated conditions (25°C/65% RH and 40°C/75%RH), an increasing trend of water content is observed. This confirms the hygroscopicity of ACA-1-HT-dry form and shows that the selected packaging is not efficient enough to prevent the water uptake. Therefore, we decided to modify the packaging using two Aluminum bags thermo-sealed under vacuum and placing between them a desiccant of silica gel. The stability studies on the new packaging were carried out in the same conditions and are reported in Table 2.3. XRPD and FTIR data confirm the stability of ACA-1-HT-dry form at different temperature and humidity conditions. Water content displays oscillations due to sample manipulation before the analysis, but it is always close to 1% and there is no an increasing trend like in the previous test. The new packaging is efficient in preventing absorption of water by Lenvatinib in all storage conditions.

|		Storage conditions and time (months)						
		5±3°C			25±2°C/65±5%R.H.		40±2°C/75±5%R.H.	
	0	3	6	12	3	6	3	6
XRPD identification	ACA-1-HT-dry	No change	No change	No change	No change	No change	No change	No change
FTIR identification	ACA-1-HT-dry	No change	No change	No change	No change	No change	No change	No change
Water content (K.F.)	1.04	0.68	0.75	1.53	1.21	0.94	1.10	0.72

Table 2.3. Stability test of Lenvatinib mesylate - ACA-1-HT-dry form in amber glass under nitrogen + 2 multilayer bag thermo-sealed under vacuum and silica desiccant.

In conclusion, the new solid form (ACA-1-HT-dry) of Lenvatinib mesylate when packed correctly is highly stable in different storage conditions.

2.4 Micronization of ACA-1-HT-dry form

Lenvatinib is administered with a dose of 4 or 10 mg and hence a low particle size is necessary for a correct dosing of the capsules. One method to reduce the particle size is micronization. However, the micronization process can perturb the solid form and it has been reported in the literature the conversion from crystalline to amorphous Lenvatinib mesylate with only 12-26% of remaining crystallinity during the formulation process.[8] Amorphization would represent an intellectual property issue because we would infringe patent US 7550483[2].

Micronization of an active pharmaceutical ingredient (API) can be achieved by different methods. We have used a jet mill, which is sketched in Figure 2.2. The solid is introduced through a funnel to a short cylinder where the micronization occurs. In this cylinder air or an inert gas is introduced through tangential nozzles creating a vortex. The solid particles collide until their dimensions are small enough to reach the center of the cylinder and go down to the collection bottle. The smallest particles can escape in the upper part of the cylinder where another collection container is placed.
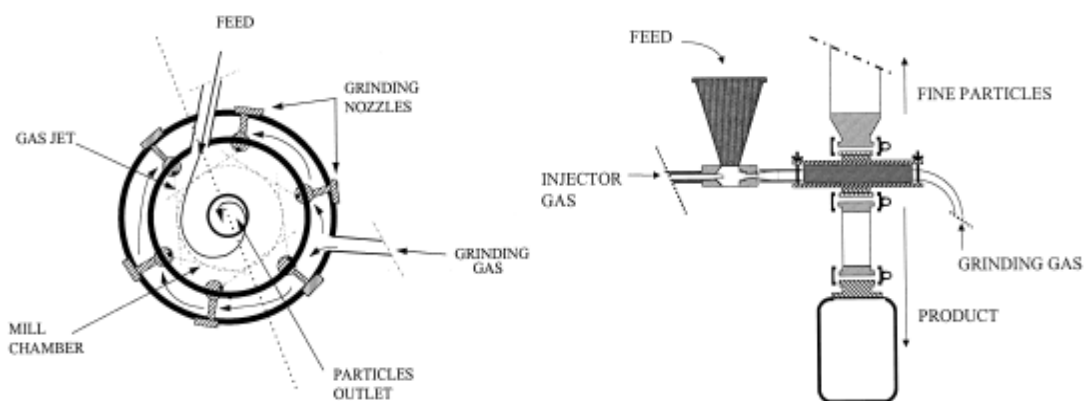
Figure 2.2. Schematic representation of jet mill micronization. (Figure modified from Reference 9)

For the micronization of Lenvatinib mesylate we used nitrogen, to avoid the presence of water which can be absorbed by the solid and can cause a solid form transition. Moreover, the feed rate was set to 10 g/min per unit of volume (L) of the micronization chamber with a feeding pressure of 5 bar, while the pressure of micronization gas was set to 6 bar. Once the feeding of the solid was completed, micronization gas was fed at the same pressure for additional 10 minutes and then the solid was recovered. The obtained solid has lower particle size (d10 = 0.6 μm, d50 = 2.3 μm, d90 = 7.7 μm) than the starting material (d10 = 3.0 μm, d50 = 25.4 μm, d90 = 231 μm). The solid was also analyzed by XRPD, DSC and FTIR. XRPD analysis (Figure 2.3, red diffractogram) shows a profile still crystalline suggesting that amorphization did not occur. However, the diffractogram is not completely identical to that of the starting material and especially after 18° 2Θ the diffraction peaks are different. Importantly, these peaks are not present in the other patented solid forms.
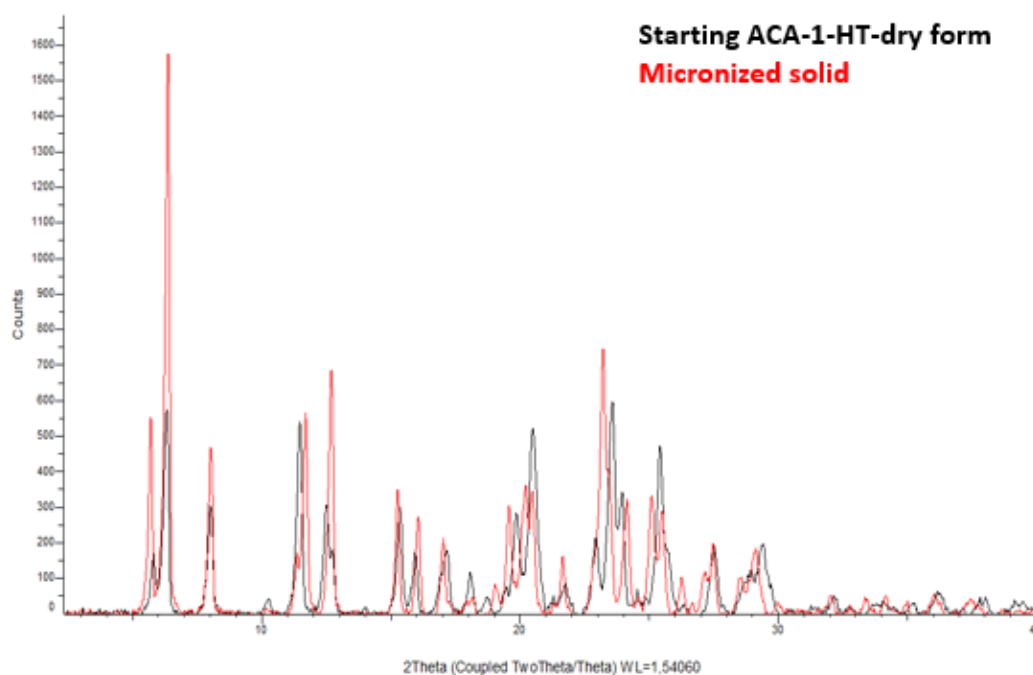
Figure 2.3. Overlay of XRPD diffractogram of starting (black) and micronized (red) solid.

DSC analysis shows a broad peak below 100°C due to release of moisture, but does not present any events associated to the presence of an amorphous phase. The melting peak has an onset at 180°C, maximum at 187°C and ΔH of about – 46 J/g. These values are very close, but not identical to DSC values of ACA-1-HT-dry form.

In conclusion, micronization of ACA-1-HT dry form does not cause conversion to an amorphous phase or any other patented solid form, but it seems to generate a new, slightly different solid form, which was covered by patent application EP 18193198.1.[10]

## 2.5 Evaluation of solubility of ACA-1-HT-dry form

In order to be approved by regulatory agencies, the generic version must have bioavailability comparable to the brand-name API. For an oral drug, this parameter is strongly correlated to its solubility, which can be affected by the solid state. Therefore, we designed preliminary experiments to assess the solubility of the new solid form (ACA-1-HT-dry) compared to the originator's polymorphs.

First of all, to have the reference materials, we prepared two anhydrous polymorphs patented by the Originator (Form A and Form C) and the amorphous. Both Form A and Form C were obtained by crystallization of Lenvatinib mesylate in methanol changing only the cooling rate and the stirring time. Indeed, to obtain Form A, the methanolic solution of Lenvatinib mesylate was cooled from 60°C to r.t slowly (5.5 h), while Form C was prepared by a faster cooling of the methanolic solution (less than 1 h) followed by stirring at r.t. for 48 h. The amorphous was obtained according to the procedure described in patent US 7550483[2]: Lenvatinib mesylate was dissolved in a mixture of water and ethanol at r.t. and the solution was concentrated at reduced pressure to remove ethanol and freeze-dried. The three solid forms were characterized by XRPD, DSC and FTIR (the details are reported in the experimental section).

The determination of the solubility was performed by suspending each solid form in a solvent and measuring the amount of Lenvatinib mesylate dissolved at the equilibrium. In order to understand when the system reached the equilibrium, we studied the kinetic of dissolution of each polymorph determining the amount of dissolved Lenvatinib by HPLC assay at different times. In these experiments it is fundamental to control that there is no change of the solid form of the analyte during the measure. Indeed, the solid can convert into a different polymorph when suspended in a solvent. We checked the stability of the polymorphs through the dissolution curves which show a discontinuity in the solubility curve if the solid form changes. Furthermore, at the end of the trial

the solid form of the insoluble residue was determined and compared to the starting one. The analysis was carried out by FTIR, which requires small amount of sample (10 mg compared to 150 mg necessary for XRPD) and is able to discriminate between the solid forms of Lenvatinib mesylate. Indeed, as shown in Figure 2.4, FTIR spectra of ACA-1-HT-dry, Form A, Form C and amorphous are different.
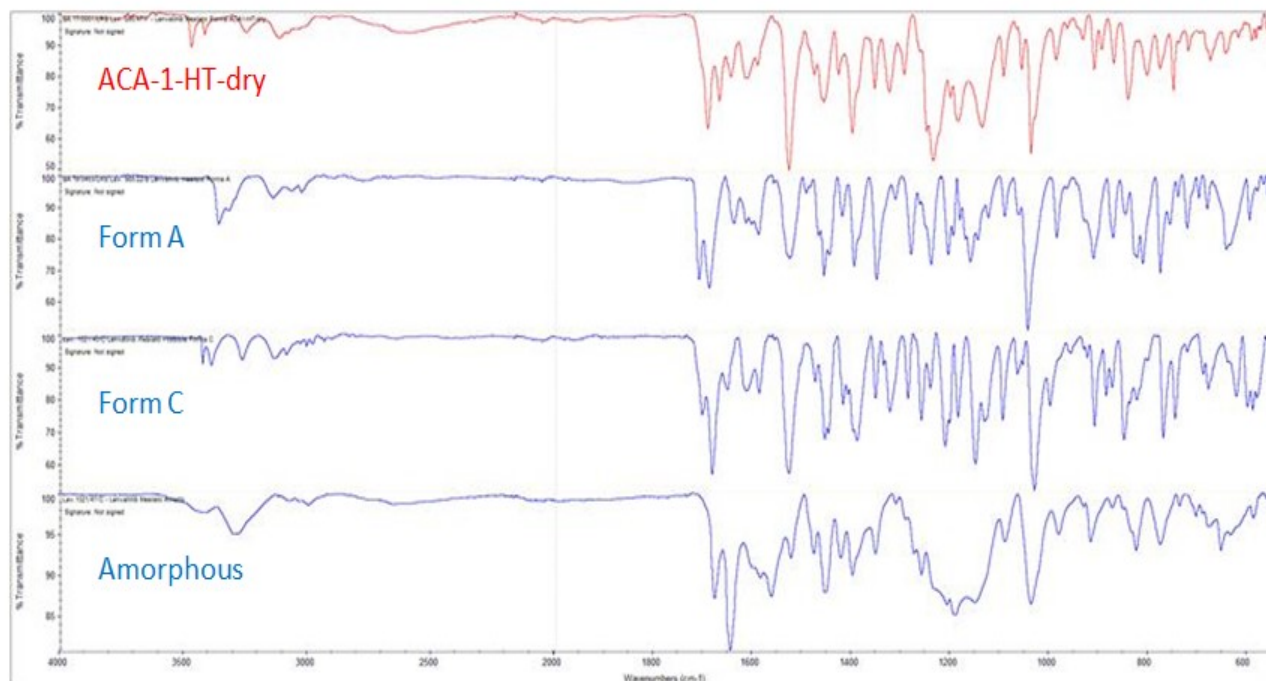
Figure 2.4. Comparison of FTIR spectra of Lenvatinib mesylate ACA-1-HT-dry, Form A, Form C, amorphous.

To mimic in vivo conditions, we tried to carry out the experiments in simulated gastrointestinal fluids, which can be prepared from commercially available powders. There are three different media: FaSSIF, which mimic fasted small intestine (pH = 6.5); FeSSIF, the simulation of small intestine at fed state (pH = 5.0); FaSSGF, which represents the fasted stomach (pH = 1.6). However, in all these simulated gastrointestinal fluids Lenvatinib mesylate forms a gel which is hard to filter and alters the measure of the dissolved amount. This behavior was detected for all the solid forms (ACA-1-HT-dry, Form A, Form C, amorphous) and in all the aqueous solution.

Therefore, we decided to run the experiments in methanol, which can be considered a suitable medium for the comparison even if not representative of in vivo fluids. Indeed, the relative solubility of the solid forms of a compound does not depend on the solvent used for the analysis. There are different energetic contributions to the overall solubility of a molecule and among them the energy necessary to break the lattice is the component responsible of the differences of solubility between polymorphs. Lattice energy does not depend on the solvent, while all the other

energies (e.g. solvation) depends strongly on the media and can be considered constant for all the solid forms of a compound dissolved in the same solvent. Thus, the experiments in methanol are not suitable to predict the absolute solubility of a polymorph of Lenvatinib mesylate in vivo, but are able to determine the relative solubility of a solid form compared to another one.

The four solid forms were suspended in methanol and the amount of dissolved Lenvatinib was determined by HPLC assay after 10, 120, 300 and 1440 min (Table 2.4 and Figure 2.5). Dissolution curve of ACA-1-HT-dry form shows a drop of solubility between 10 and 1440 min which suggests a transformation of solid from. This hypothesis was confirmed by the analysis of the insoluble residue, whose FTIR spectrum corresponds to a mixture of Form A and Form C. Similarly, the amorphous converts to mixture of Form A and Form C probably after a few minutes of suspension in methanol. On the other hand, Form A and C were stable during the analysis and Form A was found slightly more soluble than Form C.

Solid form	Solubility (mg/mL)				FTIR
	10 min	120 min	300 min	1440 min	
ACA-1-HT-dry	21.76	15.14	14.31	9.23	Conversion to A and C mixture
Form A	6.43	5.65	6.76	6.57	Stable
Form C	5.20	5.68	6.18	5.64	Stable
Amorphous	14.92	14.29	13.10	8.49	Conversion to A and C mixture

Table 2.4. Results of solubility study in methanol.
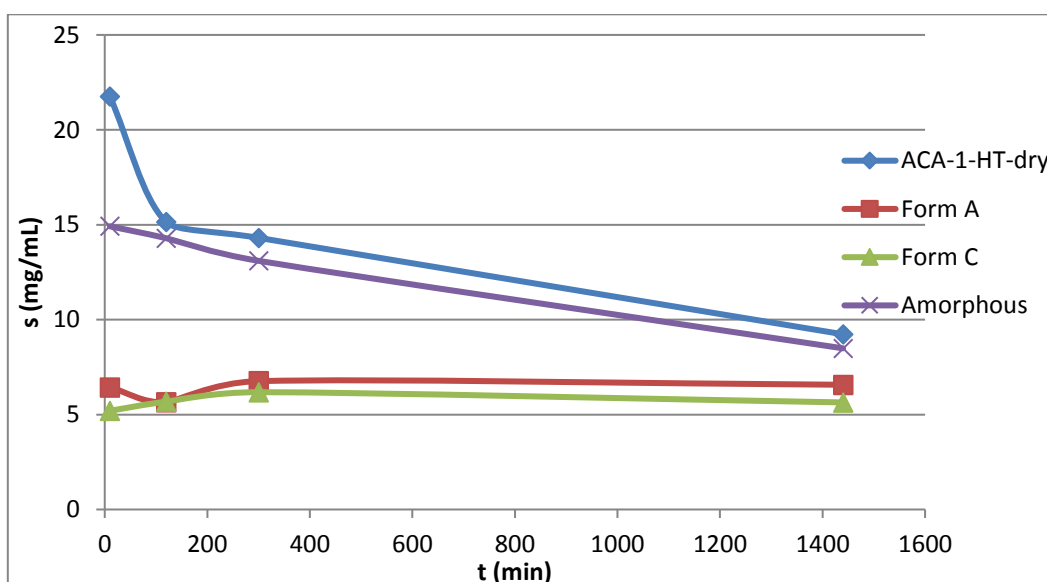

Figure 2.5. Dissolution curves of polymorphs of Lenvatinib mesylate in methanol.

Because ACA-1-HT-dry form converted into a different polymorph during the experiment, it was not possible to determine its solubility. Therefore, we decided to repeat the experiment in ethanol

where Lenvatinib mesylate is less soluble and consequently a change of solid form is less likely. Each polymorph was suspended in ethanol and the amount of dissolved Lenvatinib was determined by HPLC assay after 10, 120, 420 and 1440 min (Table 2.5 and Figure 2.6). Finally, the not solubilized solid residue was analyzed by FTIR to control the stability of the solid form. The dissolution curves and FTIR of the solid residues show that ACA-1-HT-dry, Form A and Form C were stable during the experiment, while amorphous converted to Form C probably between 10 and 120 minutes. Therefore, it is not possible to quantify exactly the relative solubility of the amorphous, but from the data obtained at 10 min (3.5 mg/mL) it seems more soluble than the crystalline forms, as expected.

Solid form	Solubility (mg/mL)				FTIR
	10 min	120 min	420 min	1440 min	
ACA-1-HT-dry	2.41	2.54	2.60	2.41	Stable
Form A	0.58	0.76	0.72	0.68	Stable
Form C	0.51	0.60	0.58	0.59	Stable
Amorphous	3.45	2.50	2.45	0.57	Conversion to Form C

Table 2.5. Results of solubility study in ethanol.
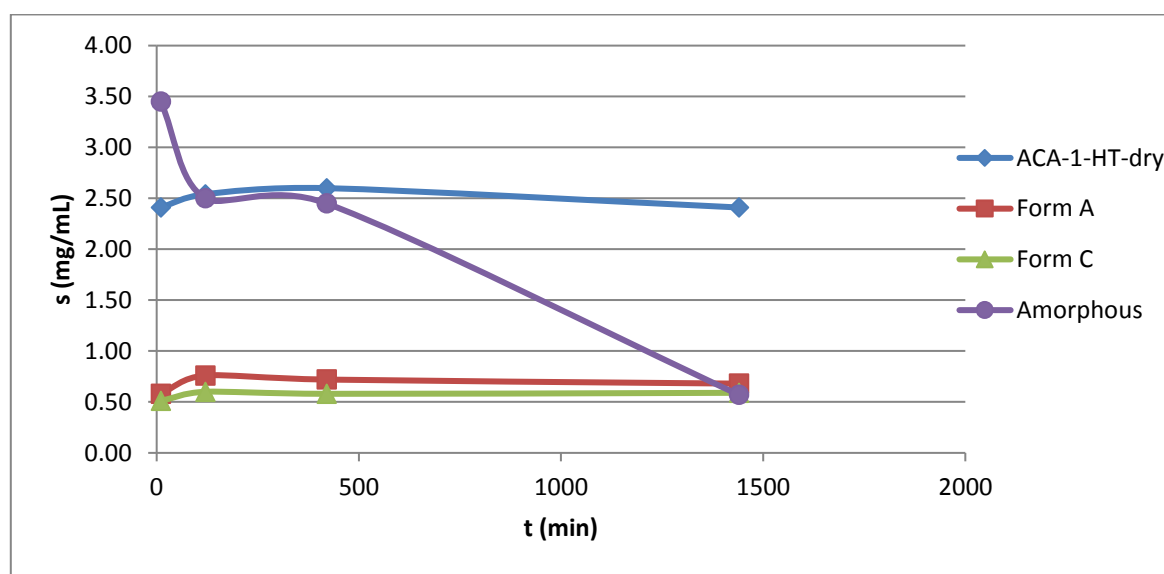

Figure 2.6. Dissolution curves of polymorphs of Lenvatinib mesylate in ethanol.

According to these data, our new polymorph, ACA-1-HT-dry, has a solubility in ethanol of 2.4-2.6 mg/mL, which is about four times higher than the values of Form A and Form C. The analysis was repeated and it was confirmed that ACA-1-HT-dry is four times more soluble than Form A and Form C.

In conclusion, we carried out the evaluation of the relative solubility of the polymorphs in ethanol where the crystalline solids do not form gels and are stable during the experiments. Our new solid form, ACA-1-HT-dry, is more soluble than the previously patented ones (Form A and Form C) and hence will not affect negatively the bioavailability of the API.

Experimental section

Methanesulfonic acid and acetic acid were purchased from Sigma Aldrich – Merck, while the other solvents (EtOAc, MeOH) were technical grade solvents commonly used in Indena plants.

XRPD diffraction patterns were recorded with a Bruker D2-Phaser diffractometer with the following conditions:

- Tube anode : Cu
- Generator tension (kV) : 30
- Generator current (mA) : 10
- Wavelengths α_1 and α_2 (Å) : 1.54056, 1.54439
- Intensity ratio (α_2/α_1) : 0.500
- Spinner : off
- Angular range ($2\theta°$) : 2.00 - 50.00
- Step size ($2\theta°$) : 0.020
- Time per step (sec) : 3.0

DSC analysis was performed using a Mettler DSC1 System. Heat flow was recorded from 30 to 250°C with linear heating rate (10°C/min), using closed aluminum crucibles (40 µl volume) with a pinhole, under a 50 ml/min nitrogen flow. About 5 mg of powder were used for each measurement. The thermal profile was acquired and elaborated by Mettler software STARe Thermal Analysis System.

DVS analysis was performed using the DVS Intrinsic1 system (Surface Measurement Systems Ltd UK). The DVS curves were acquired by Intrinsic Control Software and elaborated by the software DVS Analysis Suite. The analysis was run at a fixed temperature of 25 ± 0.1°C. The sample was dried for 6 hours under a continuous flow of dry air (Relative Humidity, RH < 0.1 %) to establish the dry mass (preconditioning step). The relative humidity was then increased from 0% to 90% RH (in 10% RH steps) and then decreased to 0% RH in a similar way, until the completion of two full sorption/desorption cycles. The instrument was run in a dm/dt mode (mass variation over time variation) and a fixed dm/dt value of 0.002%/min was selected, in order to let equilibrium to be reached at each step. A maximum dm/dt stage time of 3 hours along with minimum dm/dt stability duration of one hour were chosen. In the last step the samples were kept for 3 hours under a flow of dry air necessary for the weight equilibration.

The infrared spectrum was recorded in Attenuated Total Reflectance (ATR) mode using Fourier-Transform spectrometer Nicolet iS10 Thermo Scientific, equipped with Specac ATR Golden Gate and OMNIC software (Thermo Scientific). The spectrum was the result of the acquisition and transformation of 32 co-added scans in the 4000-550 cm^{-1} spectral region at a resolution of 4 cm^{-1}.

The ^{13}C CPMAS spectrum was acquired with a Jeol ECZR 600 instrument, operating at 600.17 and 150.91 MHz, respectively for ^{1}H and ^{13}C nuclei. The powder sample was packed into a cylindrical zirconia rotor with a 3.2 mm o.d. and 60 μL volume. A certain amount of sample was collected from the batch and used without further preparations to fill the rotor. The ^{13}C CPMAS spectrum was acquired at a spinning speed of 20 kHz at room temperature, using a ramp cross-polarization pulse sequence with a 90° ^{1}H pulse of 2.1 μs and a contact time of 3.5 ms. An optimized recycle delay of 1.7 s was used, for a number of scans of 1500. A two-pulse phase modulation (TPPM) decoupling scheme was used, with a radiofrequency field of 108.5 kHz. The ^{13}C chemical shift scale was calibrated through the methylenic signal of external standard glycine (at 43.7 ppm).

HPLC system consists of Quaternary pump, thermostated cooling autosampler, thermostated column compartment and UV/VIS detector connected to Empower software.

Stationary phase: Phenyl-Hexyl.

Mobile phase: 0.012% TFA in Water and 0.012% TFA in acetonitrile (linear gradient).

Other details of HPLC methods (e.g. flow rate, detection, injection volume, Column temperature) are not reported here for confidentiality reasons.

Preparation and characterization of Lenvatinib mesylate (ACA-1-HT-dry form)

Lenvatinib mesylate (1.5 g, 2.87) was suspended in acetic acid (4.5 mL, 3 V). The mixture was stirred at 60°C for 30' to reach complete dissolution. The solution was passed through Whatman 0.2 μm filter and heated again at 60°C. Ethyl acetate (1.5 mL, 1 V) was added drop-wise and the external temperature set at 40°C. The mixture was stirred at 40°C for 16h (at this temperature a white solid precipitates) and at room temperature for 1h. The suspension was filtered and the solid washed with 3:1 Acetic Acid : EtOAc (1.5 mL, 1V). The wet solid (ACA-1 form, 3.05 g) was dried at 70°C in vacuum for 72 h to afford the title compound (1.11 g, 2.12 mmol, y = 74 %).

XRPD: The most intense diffraction peaks (2θ angle) are the following: 5.9 – 6.4 – 8.1 – 11.6 – 12.7 – 15.4 – 16.1 – 17.2 – 18.2 – 19.9 – 20.6 – 21.8 – 23.6 – 24.1 – 25.5 – 27.5 – 29.3.

DSC:
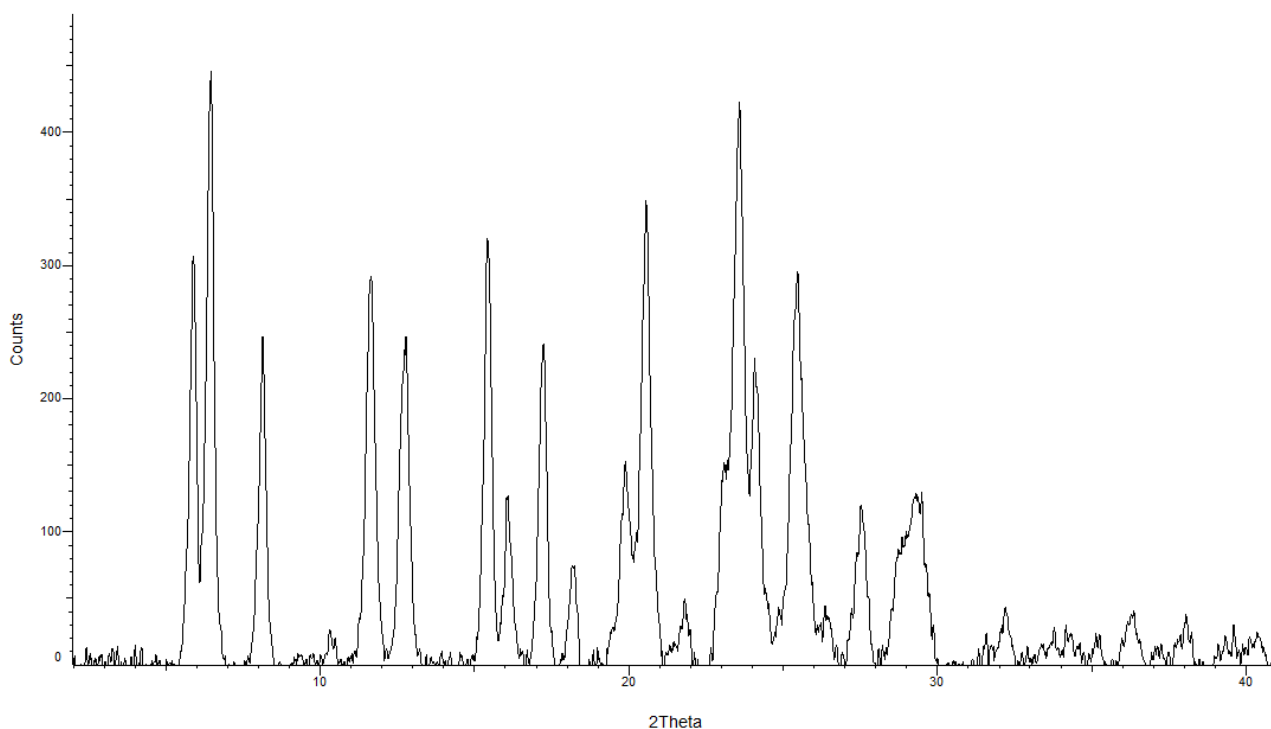
- Broad endothermic signal below 100°C due to release of residual moisture

- Melting peak with onset at 176.2°C, maximum at 189.1°C and ΔH of about - 43 J/g.

- Decomposition takes place upon melting, at about 190-200°C.
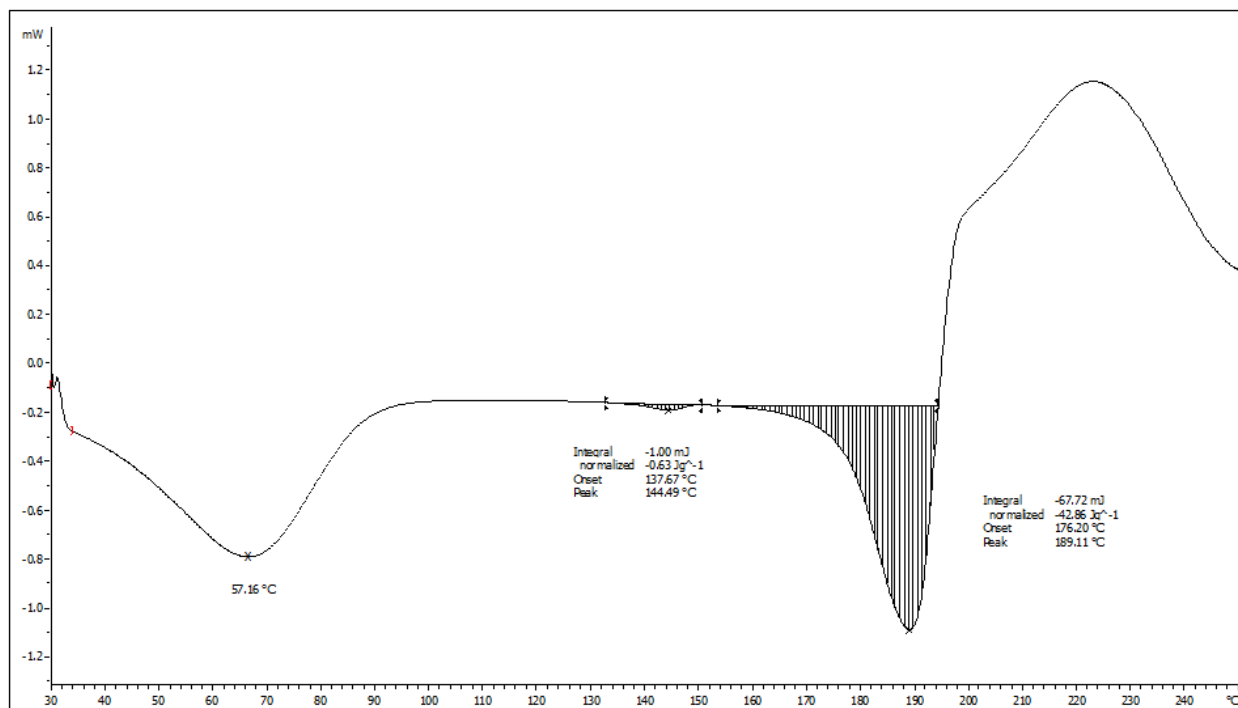
DVS: *Cycle 1*

Target RH (%)	Change in mass (%)		
	Sorption	Desorption	Hysteresis
0.0	0.00	3.22	
10.0	1.21	4.57	3.35
20.0	6.03	5.83	-0.19
30.0	6.36	6.20	-0.16
40.0	6.56	6.66	0.10
50.0	6.68	7.15	0.47
60.0	6.94	7.80	0.86
70.0	7.37	9.35	1.97
80.0	8.25	10.53	2.28
90.0	13.63	13.63	

Cycle 2

Target RH (%)	Change in mass (%)		
	Sorption	Desorption	Hysteresis
0.0	3.22	4.75	
10.0	4.42	5.60	1.18
20.0	5.66	5.92	0.26
30.0	5.99	6.19	0.20
40.0	6.31	6.45	0.14
50.0	6.68	6.76	0.08
60.0	7.16	7.16	0.00
70.0	7.78	7.69	-0.09
80.0	8.58	8.44	-0.14
90.0	10.05	10.05	
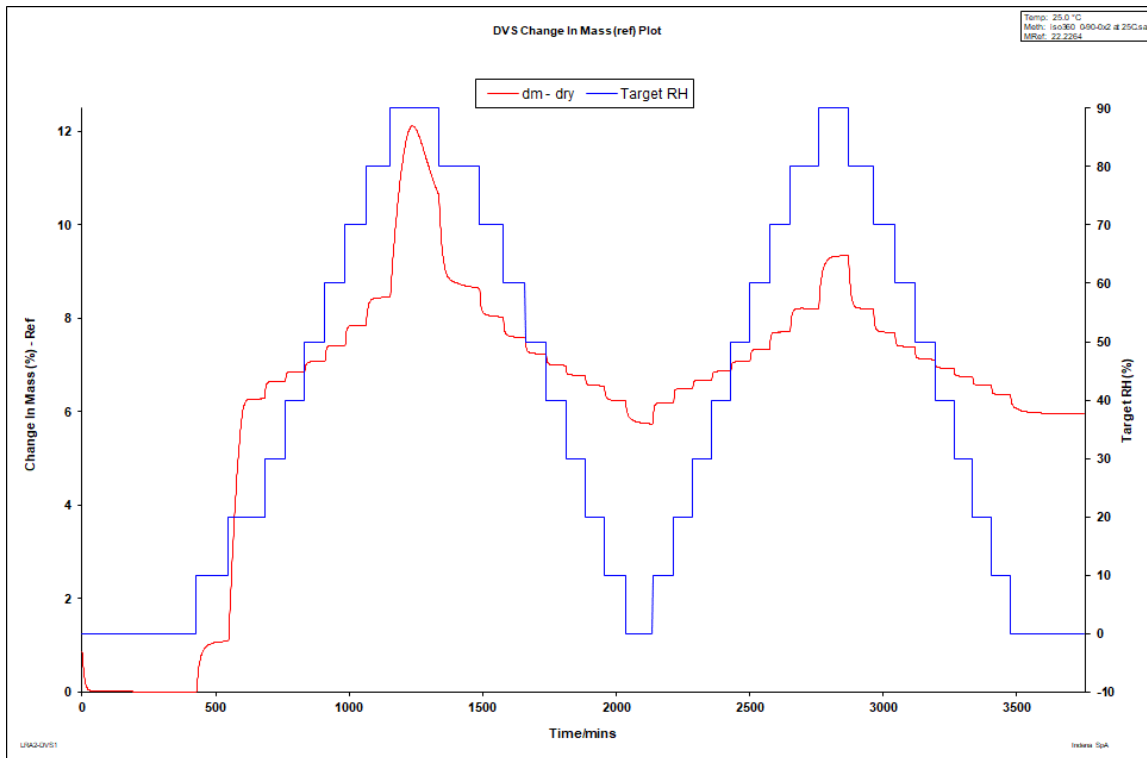
<u>FTIR</u> The most intense absorption bands are the following: 3469 – 3424 – 3247 – 3119 – 3028 – 1691 – 1668 – 1648 – 1526 – 1456 – 1399 – 1353 – 1323 – 1294 – 1240 – 1185 – 1142 – 1092 – 1055 – 1027 – 987 – 908 – 841 – 802 – 768 – 749 ± 2 cm^{-1}.
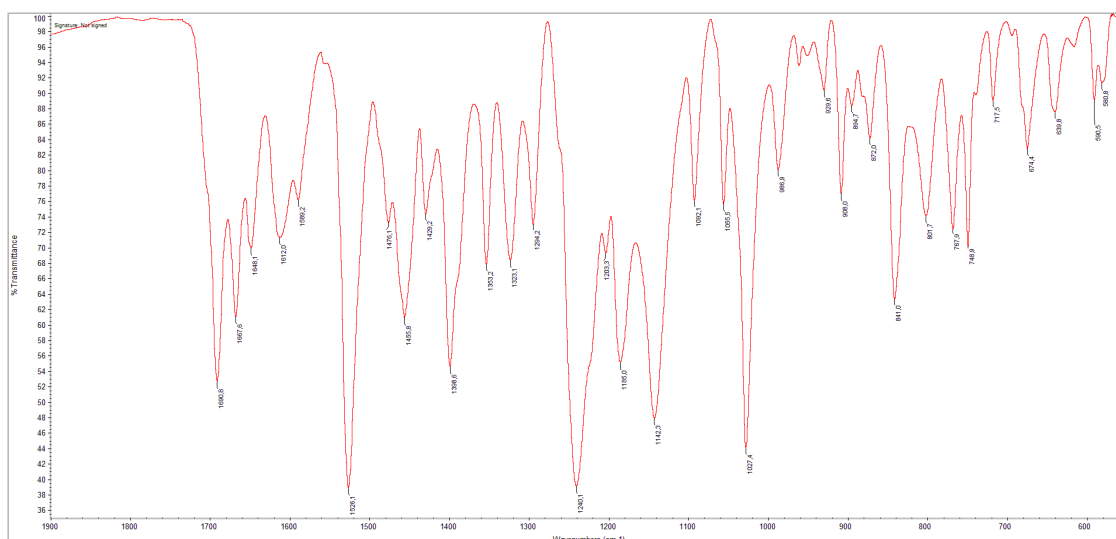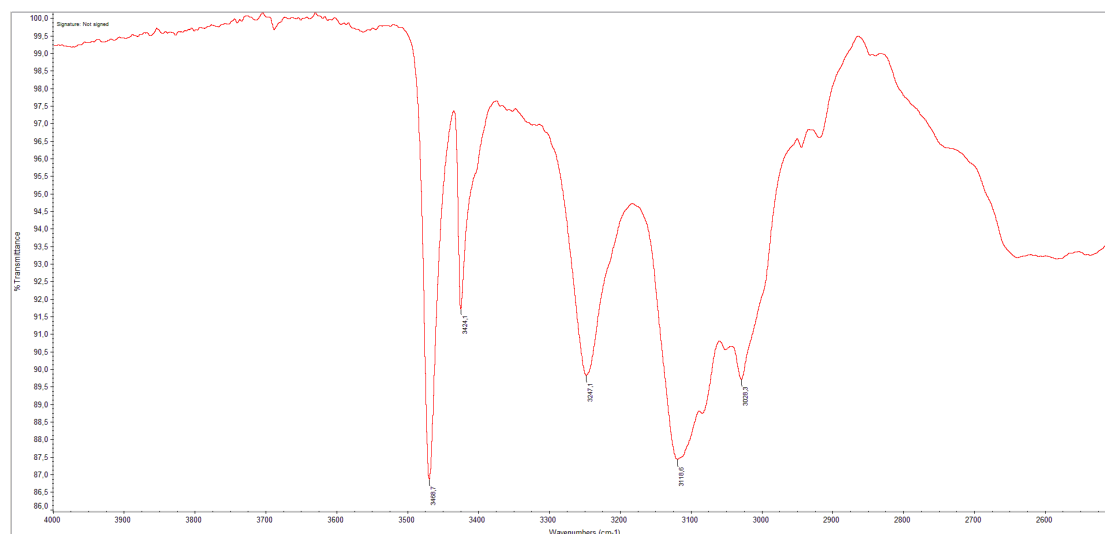
<u>^{13}C MAS NMR</u> δ (ppm): 7.9, 8.2, 22.0, 40.8 (MeSO$_3$H), 58.1, 99.8, 100.7, 106.2, 114.0, 121.1, 123.3, 129.7, 135.5, 141.4, 145.4, 146.5, 156.4, 161.2, 164.3, 169.2
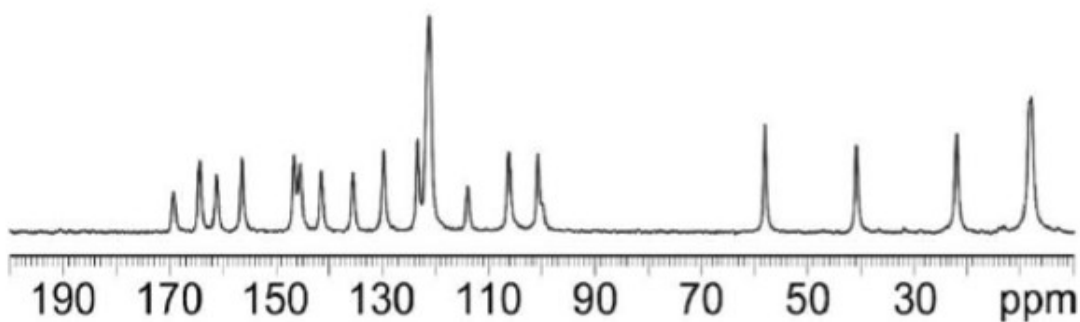
Particle Size distribution

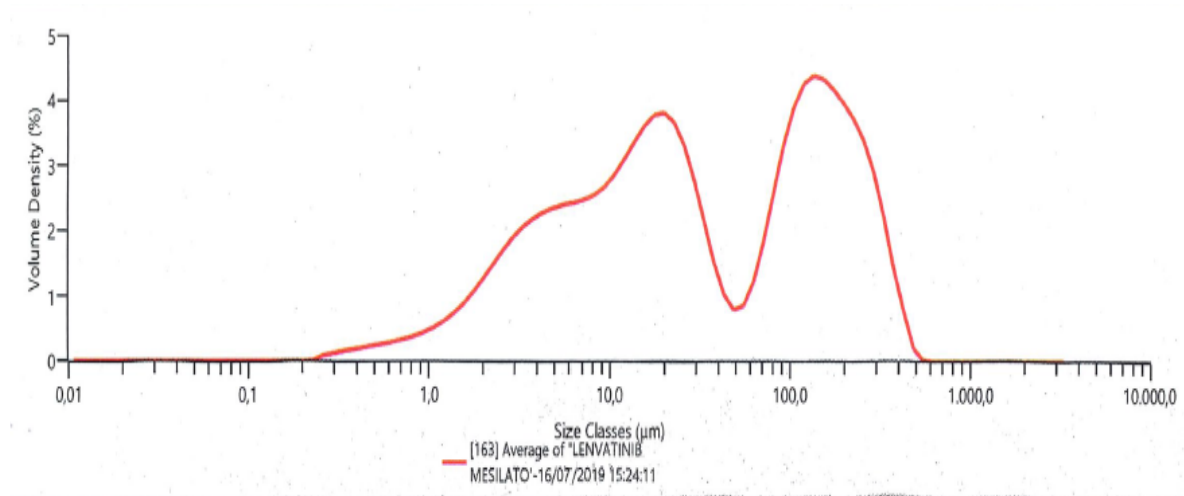d10 = 3.0 µm, d50 = 25.4 µm, d90 = 231 µm

Micronization of Lenvatinib mesylate (ACA-1-HT-dry form)

Lenvatinib mesylate, ACA-1-HT dry form (10.0 g) was micronized in MICRONET 100. The sample was introduced portion-wise in 10' (feed rate 10 g/min per volume unit (L)). Nitrogen was flown with alimentation pressure set at 5 bar and micronization pressure set at 6 bar.

After the addition was complete the pressure of alimentation was brought to 0 bar, while micronization pressure was maintained to 6 bar for 10'. Finally, the solid was recovered from collection container (7.3 g, recovery = 73%).

Particle Size distribution

d10 = 0.6 µm, d50 = 2.3 µm, d90 = 7.7 µm
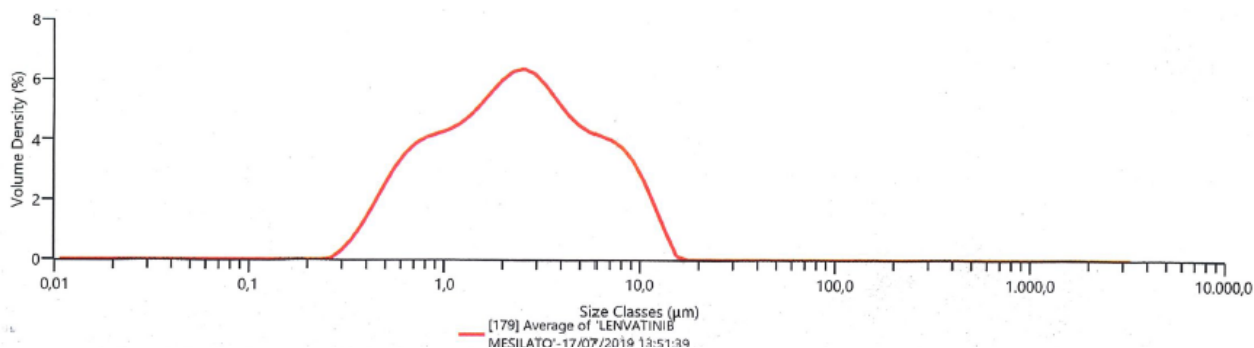
XRPD: The main peaks are at the following 2θ angle: 5.7 – 6.3 – 8.0 – 11.7 – 12.7 – 15.3 – 16.0 – 17.0 – 19.1 – 19.6 – 20.2 – 20.4 – 23.2 – 24.1 – 25.1 – 25.5 – 26.3 – 27.5 – 29.1.

40

DSC:
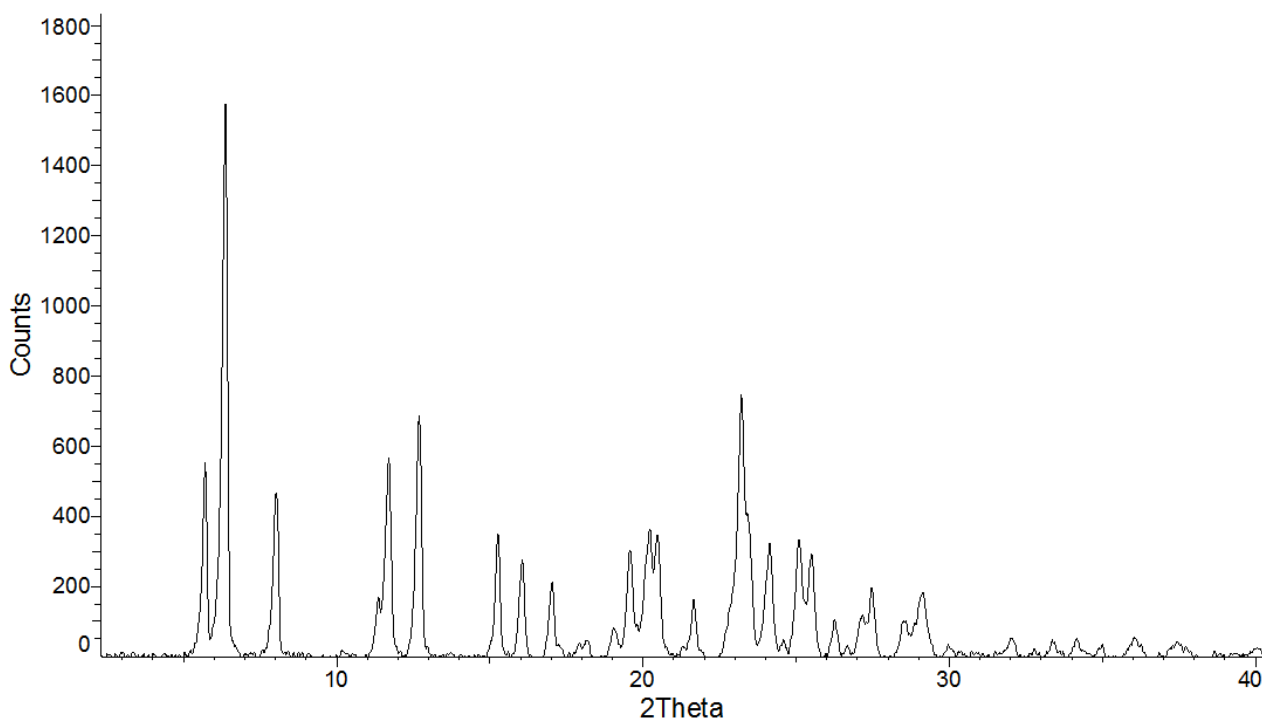
- broad endothermic signal below 100°C due to release of residual solvent/moisture
- melting peak with onset at about 180°C, maximum at about 187°C and ΔH of about - 46 J/g
- Decomposition takes place upon melting, at about 190-200°C.
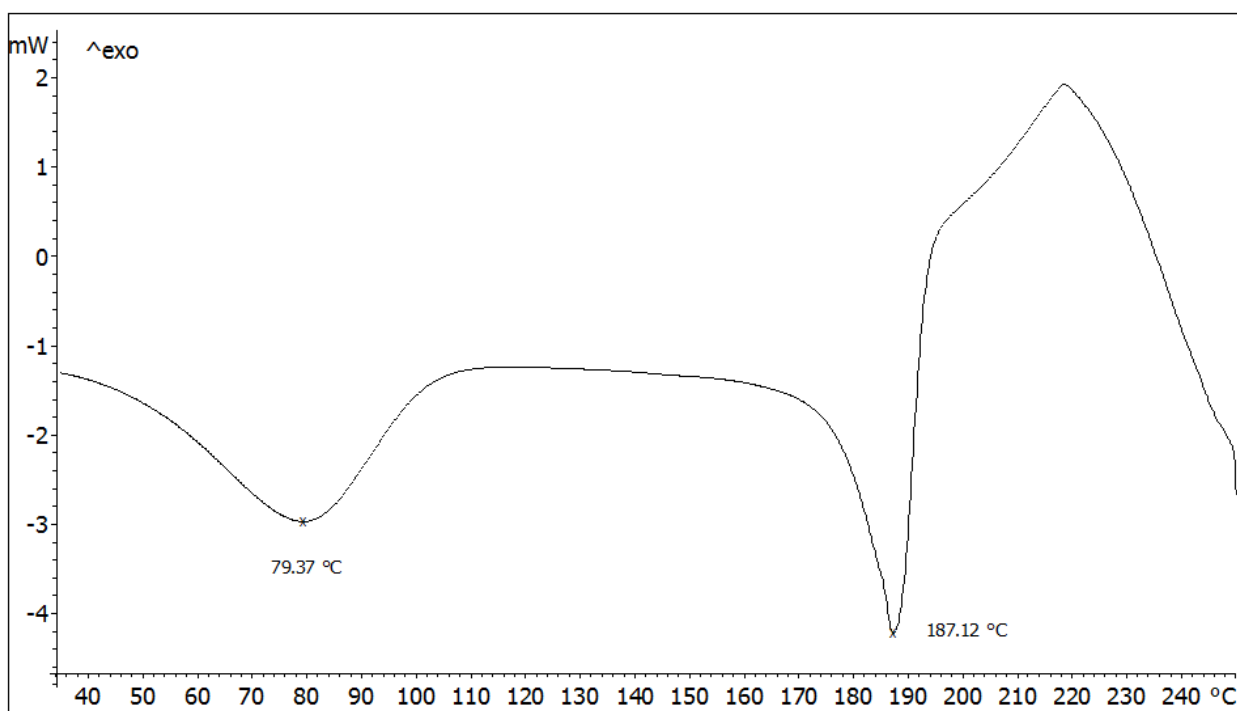
FTIR: the most intense absorption bands are the following: 3468 – 3415 – 3247 – 3113 – 1690 – 1667 – 1525 – 1455 – 1398 – 1353 – 1324 – 1234 – 1185 – 1141 – 1028 – 840 – 749 ± 2 cm-1.
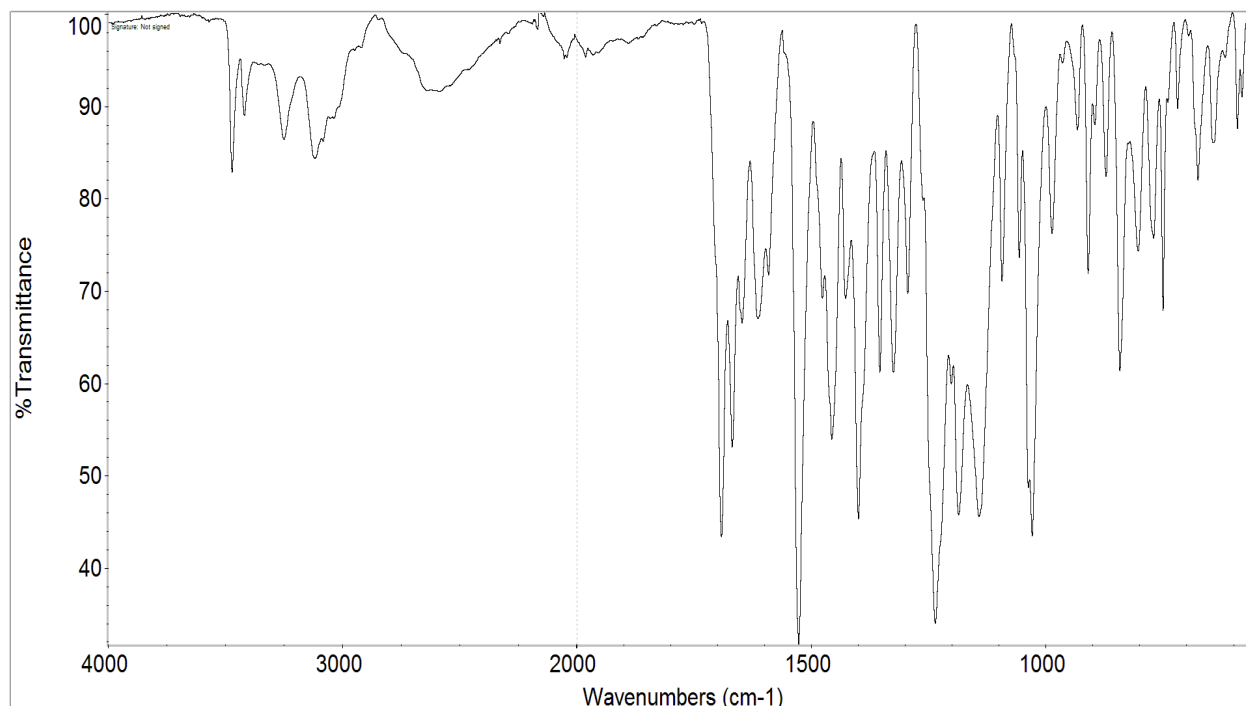
Preparation of Form A of Lenvatinib mesylate

8.0 g (18.7 mmol) of Lenvatinib free base were suspended in 136 mL (17 V) of MeOH and the suspension was heated at 60°C. Methanesulfonic acid (1.8 g, 18.7 mmol, 1 eq.) diluted in 24 mL (3 V) of MeOH was added drop-wise. At the end the addition, the mixture was cooled to room temperature in 5.5 h and stirred at room temperature for 18 h. The suspension was filtered and the solid was washed with 2 x 16 mL (2 V) of MeOH. Wet solid (10.3 g) was dried under vacuum at 60°C for 24 h to yield Lenvatinib mesylate - Form A (7.1 g, 13.6 mmol, y = 73%).

XRPD: The main peaks are at the following 2θ angle: 6.5 – 9.7 – 10.6 – 11.4 – 12.6 – 13.1 – 14.5 – 16.7 - 17.3 – 17.7 – 18.4 – 18.9 – 19.4 -20.0 – 20.4- 20.8 – 21.4 – 22.2 – 22.9 – 23.2 – 23.4 – 24.0 – 24.8 - 25.5 - 26.4 – 26.8 – 27.6 – 29.7 – 30.0 – 32.7 – 35.4.

DSC:

- melting peak with onset at about 244°C, maximum at about 248°C and ΔH of about - 33 J/g
- Decomposition takes place upon melting, at about 250°C.

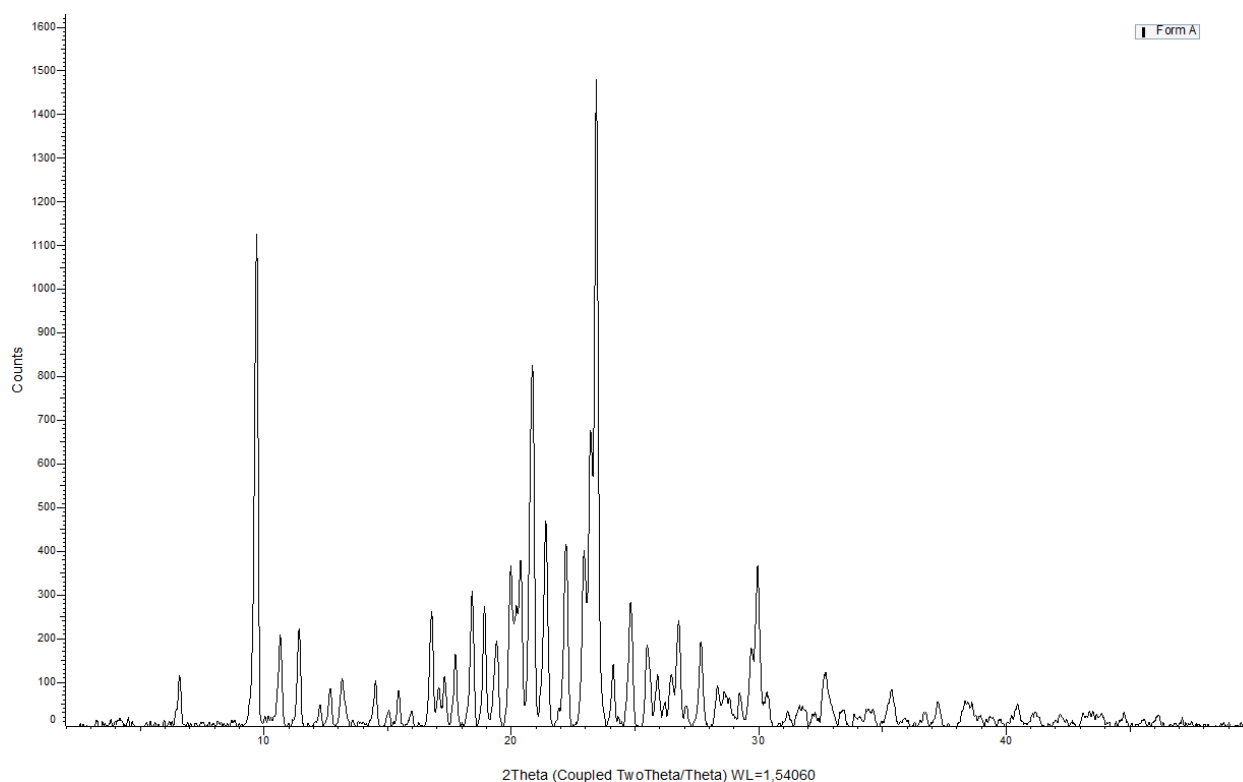FTIR: the most intense absorption bands are the following: 3306 – 3143 – 2676 – 2179 – 1709 – 1689 – 1639 – 1589 – 1526- 1492 – 1456 – 1420 – 1350 – 1311 – 1280 – 1240 – 1204 – 1194 – 1181 – 1161 – 1091 – 1044 – 985 – 911 – 846 – 827 – 811 – 756 ± 2 cm-1.
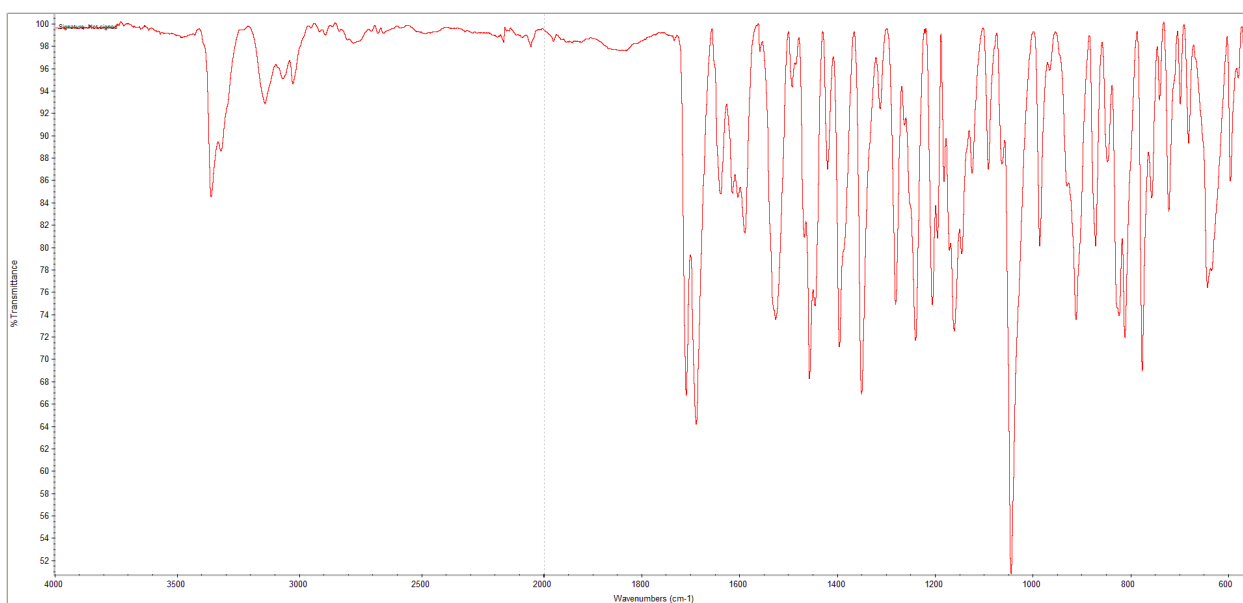
Preparation of Form C of Lenvatinib mesylate

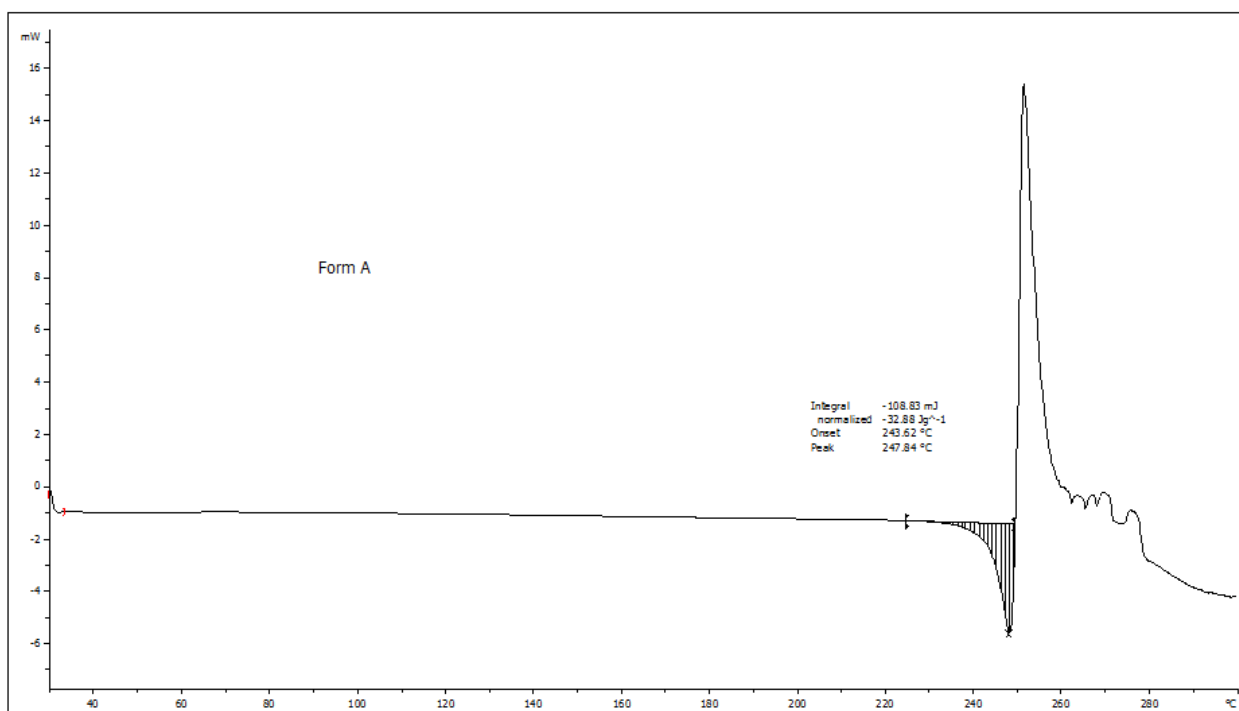5.0 g (11.7 mmol) of Lenvatinib free base were suspended in 85 mL (17 V) of MeOH and the suspension was heated at 60°C. Methanesulfonic acid (1.1 g, 11.9 mmol, 1 eq.) diluted in 15 mL (3 V) of MeOH was added drop-wise. At the end the addition, the mixture was cooled to room temperature in 0.5 h and stirred at room temperature for 48 h. The suspension was filtered and the solid was washed with 2 x 10 mL (2 V) of MeOH. Wet solid (7.0 g) was dried under vacuum at 60°C for 24 h to yield Lenvatinib mesylate - Form C (3.5 g, 6.7 mmol, y = 57 %).

XRPD: The main peaks are at the following 2θ angle: 6.2 – 9.8 – 10.1 – 10.5 – 12.3 – 14.2 – 15.2 – 17.6 – 19.0 – 19.3 – 19.9 – 20.4 – 20.8 – 21.3 – 21.7 – 22.6 – 23.1 – 23.6 – 23.7- 24.0 – 24.3 – 24.7 25.5 – 28.4 – 29.5 – 29.6 – 29.8 – 30.6 – 34.5 – 37.6.

2Theta (Coupled TwoTheta/Theta) WL=1,54060

- melting peak with onset at about 237°C, maximum at about 240°C and ΔH of about - 51 J/g
- Decomposition takes place upon melting, at about 245°C.

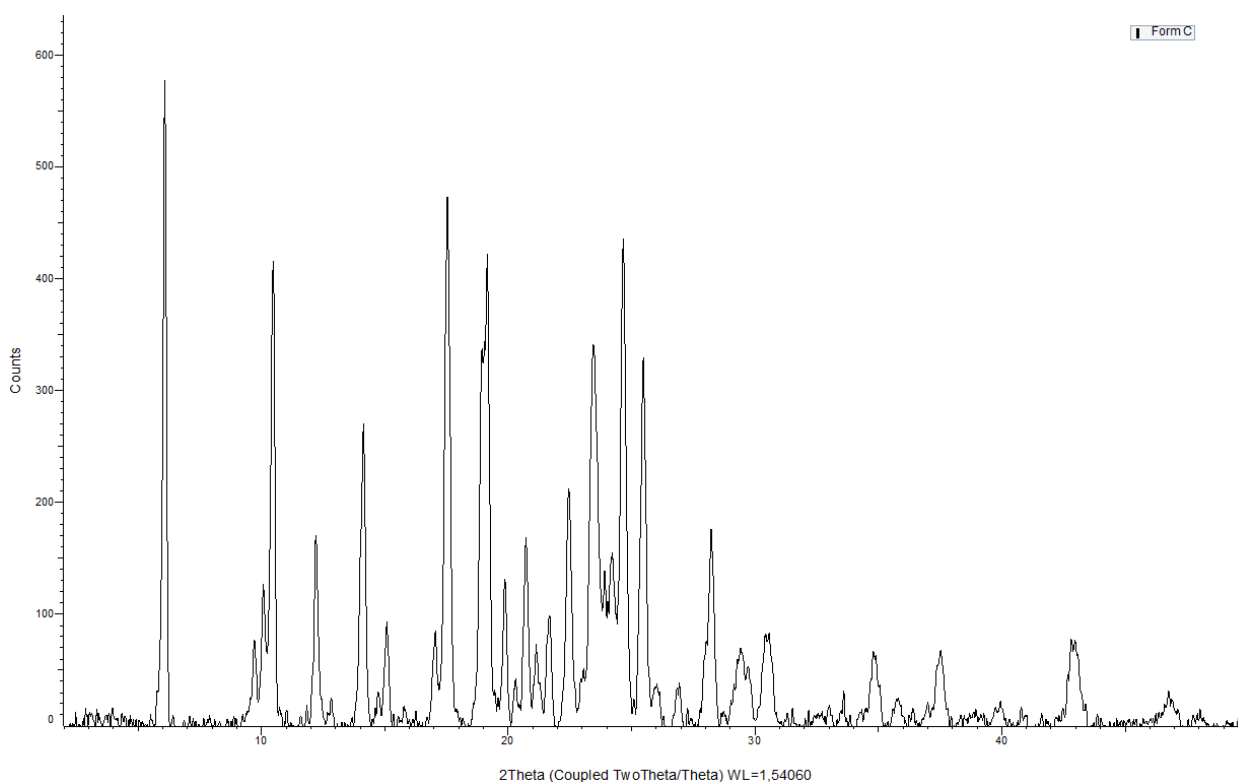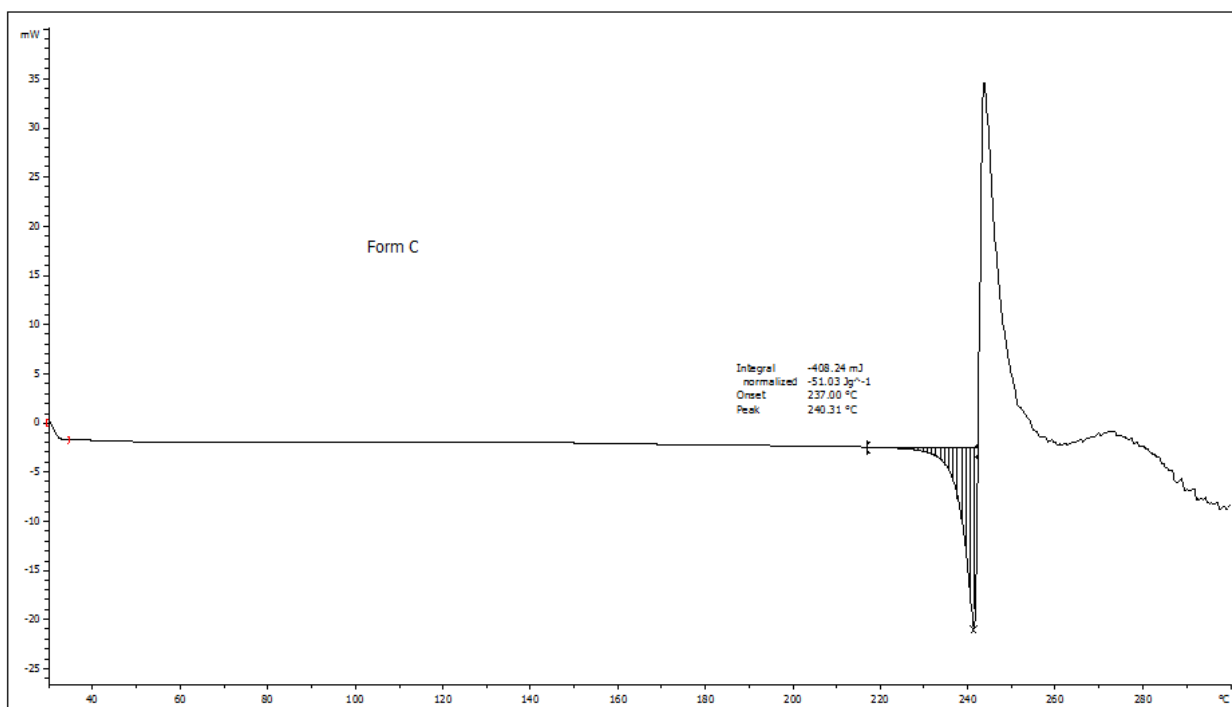FTIR: the most intense absorption bands are the following: 3423 – 3387 – 3265 – 3134 – 2189 – 2055 – 1701 – 1652 – 1613 - 1587 – 1528 – 1474 – 1417 – 1390 – 1352 – 1286 – 1241 – 1132 – 1094 – 1053 – 999 – 885 – 873 – 823 – 746 – 720 – 678 ± 2 cm-1.
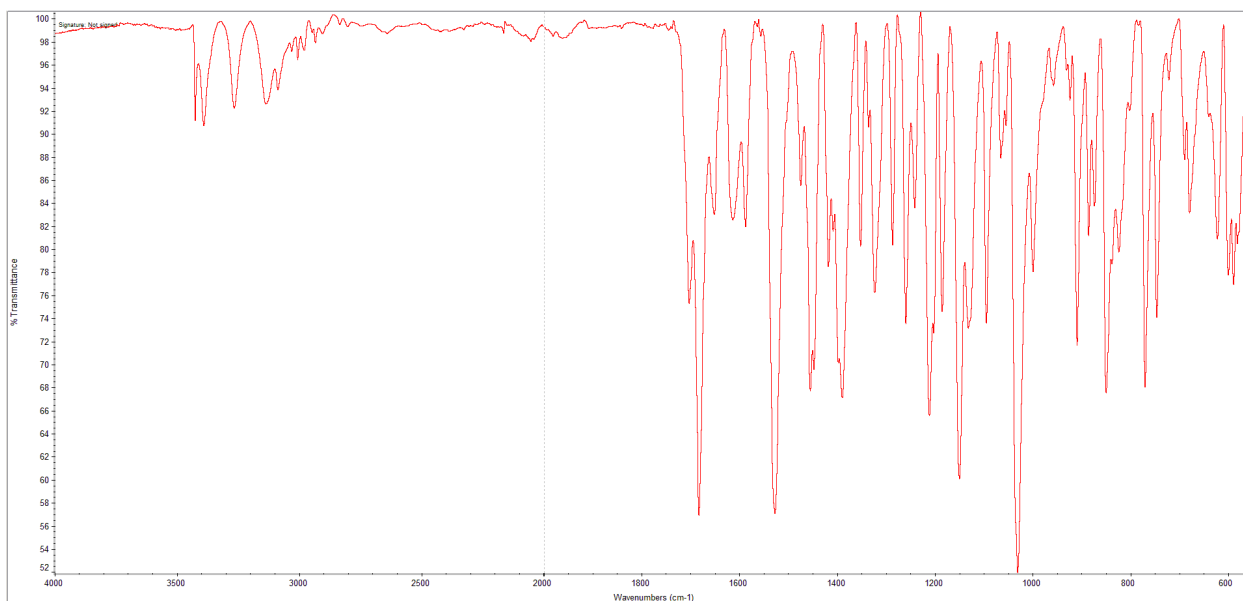
Preparation of amorphous Lenvatinib mesylate

2.0 g (3.82 mmol) of Lenvatinib mesylate (Form A) were dissolved in EtOH (80 mL, 40 V) and H_2O (80 mL, 2V). The organic solvent was removed at reduced pressure and the resulting aqueous solution was freeze-dried to afford amorphous Lenvatinib mesylate (1.90 g, 3.63 mmol, y =95%).

XRPD: absence of diffraction peaks
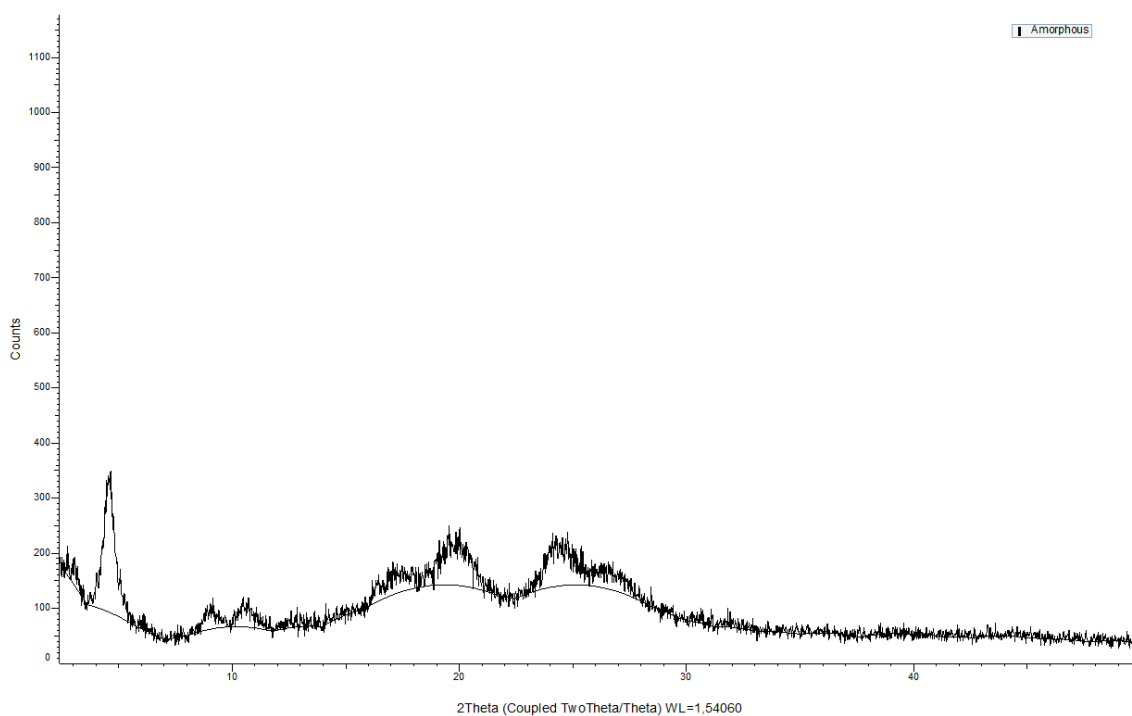
FTIR: the most intense absorption bands are the following: 3287 – 2050 – 1677 – 1644 – 1588 – 1527 – 1452 – 1421 – 1398 - 1351 – 1290 – 1238 – 1187 – 1039 – 914 – 874 – 839 – 774 – 737 – 684 ± 2 cm-1.
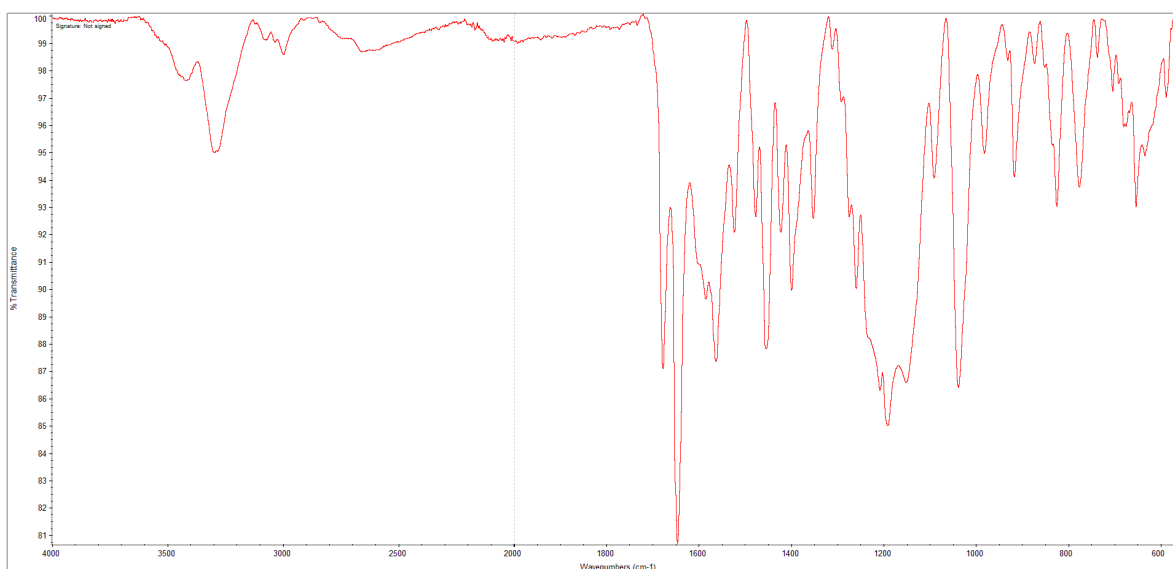
Evaluation of solubility of polymorphs in ethanol

100 mg of each solid form of Lenavtinib mesylate (ACA-1-HT-dry, Form A, Form C, amorphous) were suspended in 10 mL of EtOH. The suspensions were stirred at 22°C. The amount of dissolved Lenvatinib mesylate was determined by HPLC assay after 10, 120, 420 and 1440 min. After 1440 min the suspension was filtered and the solid was analyzed by FTIR.

Typical HPLC chromatogram
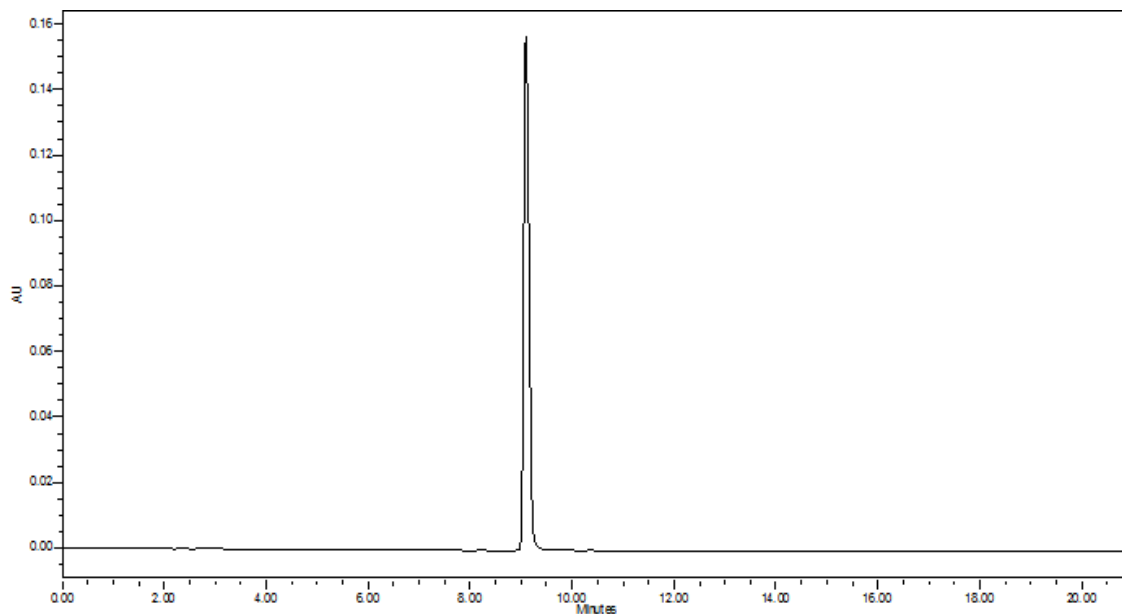
Amount of Lenvatinib dissolved

Solid form	Dissolved Lenvatinib (mg/mL)			
	10 min	120 min	420 min	1440 min
ACA-1-HT-dry	2.41	2.54	2.60	2.41
Form A	0.58	0.76	0.72	0.68
Form C	0.51	0.60	0.58	0.59
Amorphous	3.45	2.50	2.45	0.57

FTIR: We calculated the correlation of FTIR spectra of each solid residue with the spectra of the reference ACA-1-HT-dry, Form A, Form C and amorphous Lenvatinib mesylate.

Starting material	Correlation FTIR spectrum of insoluble residue				Conclusion
	ACA-1-HT_dry	Form A	Form C	Amorphous	
ACA-1-HT-dry	97.42	10.67	27.56	19.95	Stable
Form A	13.29	99.01	14.01	27.38	Stable
Form C	32.87	21.40	99.67	28.53	Stable
Amorphous	33.13	19.29	99.58	28.38	Conversion to Form C

References

[1] T. Matsushima, T. Nakamura, K. Yoshizawa, A. Kamada, Y. Ayata, N. Suzuki, I. Arimoto, T. Sakaguchi, M. Gotoda, **2004**, patent US 7612208, patent EP 1698623.

[2] T. Sakaguchi, A. Tsuruoka, **2006**, patent US 7550483.

[3] C. Minhua, Z. Yanfeng, D. Xiaojiuan, Z. Xiaoyu, **2015**, patent application WO 2016/184436.

[4] C. Minhua, Z. Yanfeng, Z. Long, Z. Po, H. Chunxiang, Y. Chaohui, Z. Xiaoyu, **2017**, patent application WO 2018/196687.

[5] V. Saladi, M. Krishna, P. Vishweshwar, **2016**, patent application WO 2018/122780.

[6] European Pharmacopeia 6.0 section 5.11, page 659.

[7] N. Sardone, S. Giaffreda, A. Gambini, A. Petrolati, P. Allegrini, E. Modena, **2016**, patent application EP 3299360.

[8] L. Lee, P. D'Angelo, D. Verbel, G. Martinez, J. Aluri, D. Brimhall, *Int. J. Clin. Pharmacol. Ther.*, **2015**, 53, 190-198.

[9] N. Midoux, P. Hosek, L. Pailleres, J. Authelin, *Powder Technol.*, **1999**, 104, 113-120.

[10] L. Senaldi, D. Ciceri, P. Allegrini, M. Ricotti, 2018, patent application EP 18193198.1.

3. Development of the process

As explained in the introduction chapter, the methods for the preparation of the API are claimed in patents US7683172[1] and US8058474[2] and are viable for the preparation of Lenvatinib mesylate. According to originator's patents, there are mainly three different paths to prepare Lenvatinib **1** (Scheme 3.1).
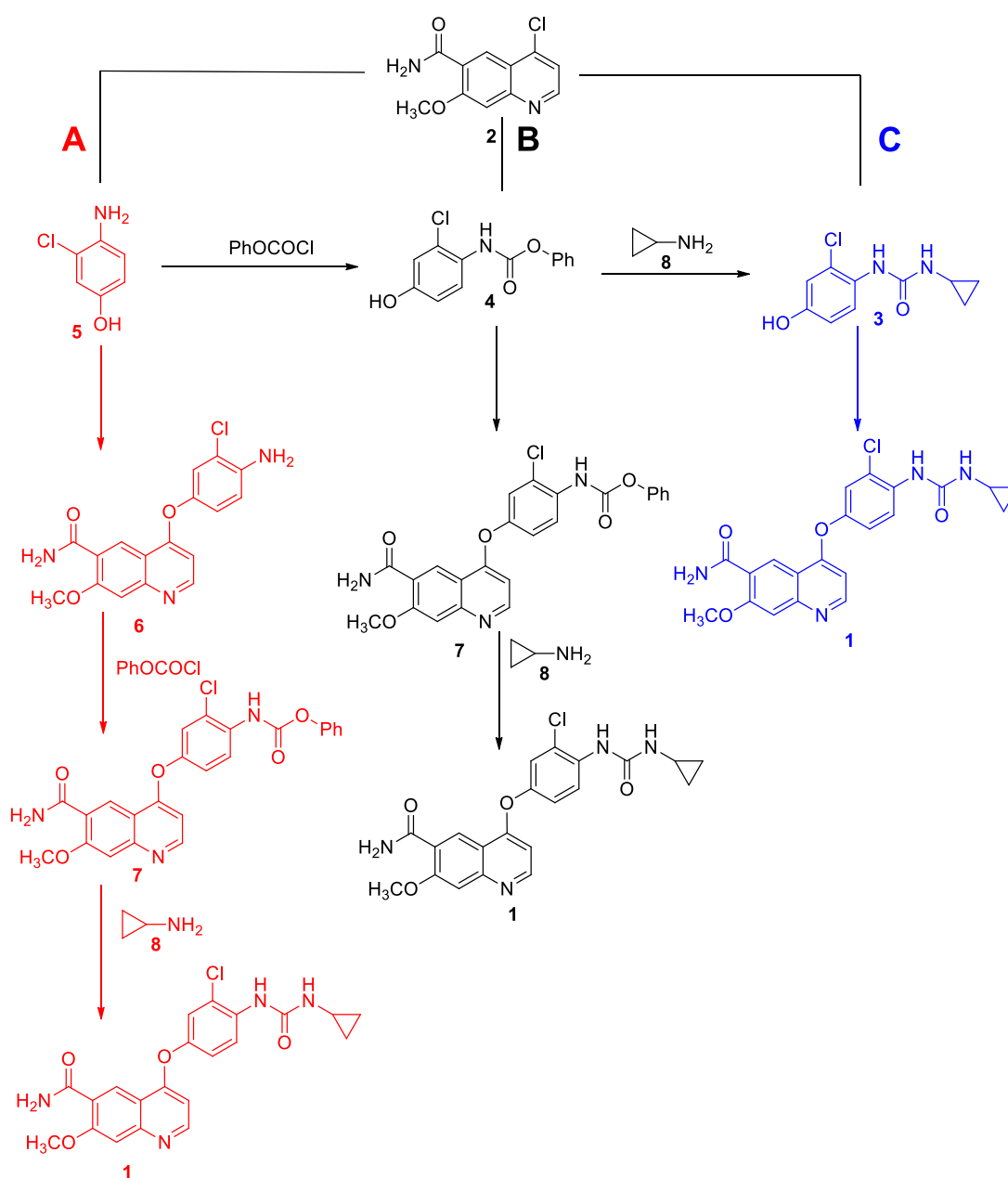
Scheme 3.1. Synthetic routes.

50

The first one, A in Scheme 3.1., starts from the coupling between 4-chloro-quinoline **2** and 4-amino-3-choro-phenol **5** and then builds the cyclopropylurea moiety via carbamate **7**. According to the second approach (B in Scheme 3.1), 4-chloro-quinoline **2** is attached to phenyl carbamate **4** followed by substitution with cyclopropylamine **8** to afford Lenvatinib **1.** In the third approach (C) quinoline **2** is directly coupled to 1-(2-chloro-4-hydroxyphenyl)-3-cyclopropylurea **3**, which is prepared from 4-amino-chlorophenol **5**. Among these strategies we focused on the optimization of synthetic route C, which was supposed to be more convergent and with higher yields according to patents. Goal of our project is to optimize the processes described in the above patents and to adopt it to industrial implementation and quality requirements. As we do not know the specifications of the API originator and the impurities that were toxicologically qualified, our generic version will have to have all the not identified impurities below 0.10% and the identified and characterized impurities below 0.15%, according to ICH Q3B guideline.[3]

3.1 Synthesis of 1-(2-chloro-4-hydroxyphenyl)-3-cyclopropylurea

According to the selected synthetic route, the first part of the process is the synthesis of 1-(2-chloro-4-hydroxyphenyl)-3-cyclopropylurea **3.** In originator's patents its preparation is described in two steps and the first one is the preparation of phenyl carbamate **4** with phenyl chloroformate, using pyridine as a base in dimethylformamide. When the reaction is complete, the mixture is diluted with ethyl acetate, washed with an acidic aqueous solution and concentrated to dryness to afford compound **4**.
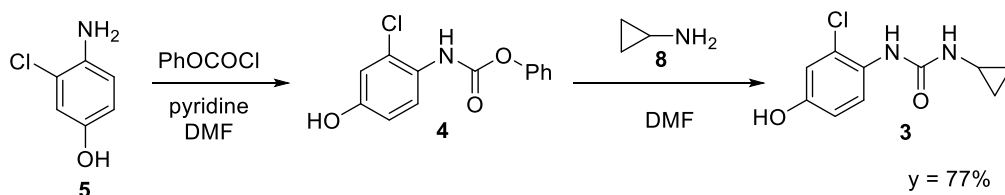
Scheme 3.2. Urea synthesis in patent US 7683172.

The hydrochloride salt of phenol **5** is commercially available, while the free amine is not present on the market. Thus, we decided to start from the hydrochloride salt of phenol **5** and we had to find an appropriate base to deprotonate the ammonium salt (Table 3.1).

Entry	Base	Solvent	Concentration (M)	PhOCOCl equivalents	Reaction time	Yield
1	Pyridine	Pyridine	1.85	1.15	40'	33%
2	Pyridine	DMF	1.65	1.15	4h	63%
3	NaHCO$_3$	THF/Water	0.2	1	15'	83%
4	NaHCO$_3$	THF/Water	0.2	1.15	15'	99%

Table 3.1. Conditions for phenyl carbamate 4 formation (experiments performed on 1g of phenol **5**)

In the first experiment (Table 3.1, Entry 1) pyridine was used both as base and as solvent and this caused partial deprotonation of the phenol group favoring formation of bis-acylated compound **9** (Figure 3.1) in about 1:1 ratio with the desired product **3**, according to the ^{1}H-NMR spectrum of the crude. The use of DMF as solvent and 1 equivalent of pyridine as base (Entry 2) decreased the amount of the byproduct, but about 21% was still present. Therefore, we changed completely the conditions with the use of a biphasic system (THF and an aqueous solution of sodium bicarbonate)[4] (Entry 3), which allowed to avoid the formation of byproduct **9**. Nonetheless, at the end of the reaction we recovered 9% of unreacted phenol **2**, so in experiment 4 an excess of phenyl-chloroformate was used (Table1, Entry 4). In these conditions the yield reached 99% and the byproduct **9** was kept below 1%.
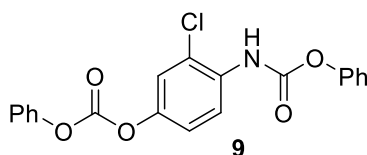
Figure 3.1. Byproduct of the reaction of formation of phenyl carbamate 4.

In the second step of the synthesis, phenyl carbamate **4** is converted into the corresponding cyclopropylurea **3** by reaction with cycloproylamine **8** in DMF at room temperature. At the end of the reaction ethyl acetate is added and the organic phase is first washed with an acidic aqueous solution and then concentrated to crystallize urea **3**. The yield calculated on the isolated product is not excellent (77% claimed in the patents and 65% in the experiments in our labs), even if the conversion is complete. Indeed, probably there is a loss of the product during the final work-up due to the presence of DMF.

Moreover, the patented synthesis requires the isolation of intermediate **4**, which is a drawback. Therefore, we decided to telescope the two steps and to avoid the use of DMF as solvent to have a more efficient process. (Scheme 3.3). The first step was performed in a biphasic system as

described before, but 2-methyltetrahydrofuran (2-MeTHF) was used as organic solvent instead of tetrahydrofuran (THF). 2-MeTHF is less miscible with water and hence it was possible to separate the aqueous phase when the conversion of the starting material to phenyl carbamate **4** was complete. The intermediate **4** was not isolated and cyclopropylamine **8** was added drop-wise directly to 2-MeTHF solution. The mixture was stirred at 50°C for 3h and heating was necessary because this step is significantly slower in 2-MeTHF than in DMF, as expected.
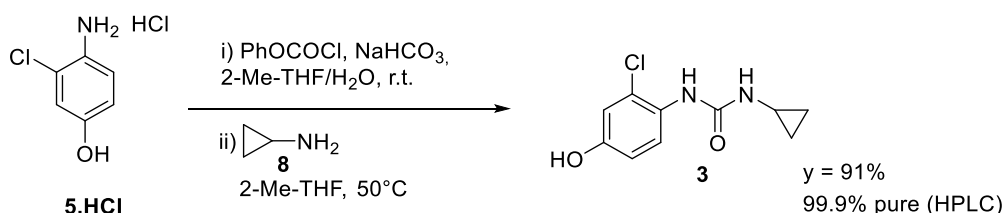
Scheme 3.3. One-pot synthesis of 1-(2-chloro-4-hydroxyphenyl)-3-cyclopropylurea 3.

When the reaction was complete, an acidic washing was carried out to remove the excess of cyclopropylamine **8**. The organic phase contains the desired product **3** and the phenol generated from cleavage of phenyl carbamate **4**. To remove the phenol, crystallization of urea **3** from the reaction mixture is required and a good crystallization solvent is ethyl acetate. In order to exchange the solvent from 2-MeTHF to EtOAc at the end of the reaction, we developed a procedure based on three distillations of the reaction mixture, followed by addition of fresh EtOAc each time. Finally, 1V of heptane is added to ethyl acetate solution (4 V) and urea **3** is obtained with 99.9% of HPLC purity and 91 % yield.

In conclusion, we developed the process for one-pot synthesis of 1-(2-chloro-4-hydroxyphenyl)-3-cyclopropylurea **3**, avoiding the isolation of phenyl carbamate intermediate **4**. Moreover, our process allows to reach excellent HPLC purity (more than 99%) and has a higher overall yield than the claimed process (90% vs 77%).

3.2 Coupling reaction

The next step of the synthesis of Lenvatinib is the coupling between 1-(2-chloro-4-hydroxyphenyl)-3-cyclopropylurea **3** and 4-chloro-7-methoxyquinoline-6-carboxamide **2** in the presence of a base. The reaction was initially performed according to the conditions reported in patent US7683172 (Scheme 3.4): 1 eq. of quinoline **2**, 1.2 eq. of urea **3** and 2 eq. of Cs_2CO_3 in DMSO were stirred for 23 h at 70 °C. When the reaction was complete, the mixture was cooled to room temperature and the product **1** was precipitated by addition of water.
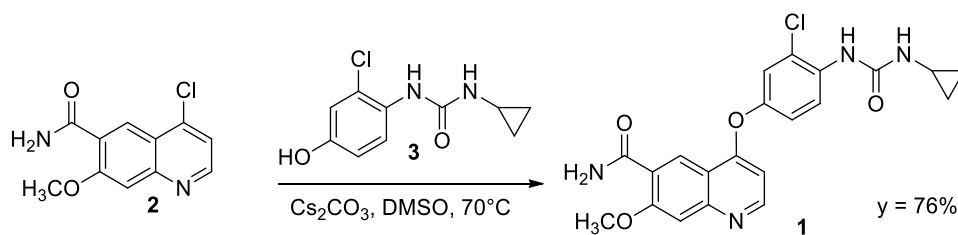
Scheme 3.4. Coupling reaction in patent US7683172.

3.2a Evaluation of the base

In the patented process cesium carbonate is used as a base to deprotonate the phenol group present in urea **3**. However, Cs_2CO_3 is rather expensive (more than 500 €/ Kg) and the variability of its physicochemical properties (crystalline form, water content) can represent an obstacle in the standardization and reproducibility of a process. Therefore, we ran preliminary experiments to evaluate the use of alternative inorganic bases, especially potassium carbonate (Table 3.2). The first experiment (Entry 2) was run in the same conditions (temperature, time and solvent) changing only the base from Cs_2CO_3 to K_2CO_3 and a clear decrease of reaction rate was observed. Indeed, after 24 h the conversion was only 48% compared to complete conversion of Entry 1. Thus, we increased the reaction temperature to 90°C (Entry 3) and the conversion increased to 85%. However, the higher temperature favored the formation of degradation products evident in the ^{1}H-NMR spectrum. The lower reactivity of K_2CO_3 compared to Cs_2CO_3 is likely due to the formation of a tight ion pair with phenate with a consequent detrimental effect on its reactivity. Therefore, we tried to run the reaction in the presence of a sub-stoichiometric amount of a quaternary ammonium salt (e.g. tetrabutylammonium hydrogensulfate) to cleave the tight ion pair, but the conversion at 70°C was not increased (Entry 4). Finally, we evaluated the use of a crown ether (18-crown-6), which is able to hijack Potassium cations in its cavity.[5] However, the conversion did not raise significantly (Entry 5).

Entry	Base	Additives	T (°C)	Time (h)	Conversion (*)	Notes
1	Cs_2CO_3	/	70°C	24	Complete	
2	K_2CO_3	/	70°C	24	48%	
3	K_2CO_3	/	70°C + 90°C	24 + 24	85%	Impurities in ^{1}H-NMR
4	K_2CO_3	TBA HSO_4^-	70°C	24	28%	Impurities in ^{1}H-NMR
5	K_2CO_3	18-crown-6	70°C	24	56%	
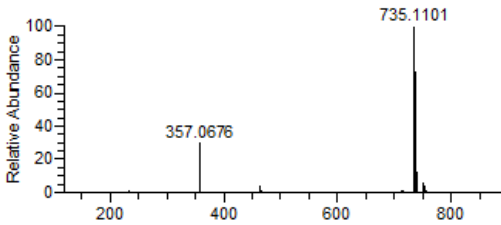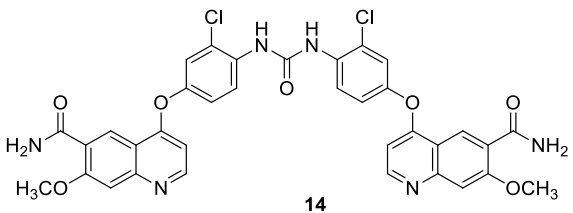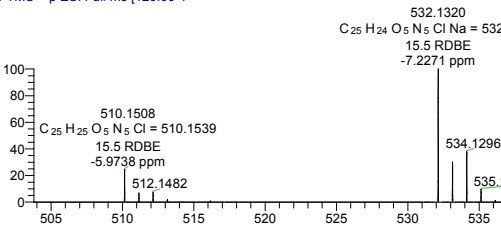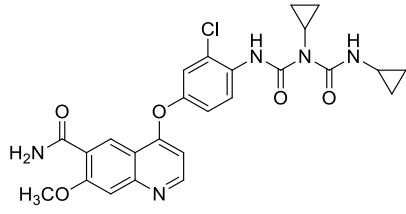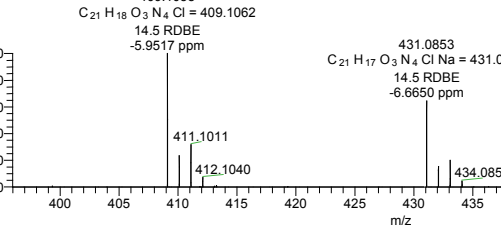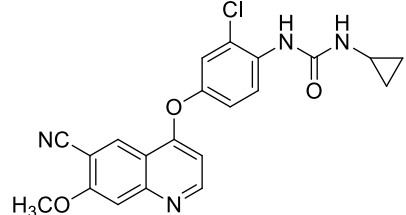
Table 3.2. Evaluation of alternative bases in coupling step. (*) determined by ^{1}H-NMR

In conclusion, Potassium carbonate does not allow to reach satisfactory conversion in reasonable reaction time and with good purity and hence we decided to maintain Cesium carbonate as the base in the coupling step.

3.2b Identification of the critical impurities

After the initial study on the base, we focused on the quality of the product obtained at the end of the coupling step. Thus, the impurities profile was studied by HPLC-MS analysis of the precipitate obtained upon coupling and precipitation performed according to the state of art (US 7683172). Typical HPLC purity of the precipitate was 95.0 % and MS and MS-MS spectra allowed to identify the structure of the impurities harder to remove (Table 3.3).

RRT	Mass spectrum	Plausible structure
0.45	CPS_HCOOH_FT_170323145140 #526-536 RT: 4.05-4.14 A F: FTMS + p ESI Full ms [100.00-1... 219.0763 $C_{11}H_{11}O_3N_2$ = 219.0764 7.5 RDBE -0.6417 ppm 264.1340	**10**
0.61	FTMS + p ESI Full ms [100.00-1000.00] 387.0853 $C_{18}H_{16}O_4N_4Cl$ = 387.0855 12.5 RDBE -0.2954 ppm	**11**
0.80	Base_1021_14_B_FTDDS #387 RT: 6.51 AV: 1 SB: 2 6.38 , 6.81 NL: 2.46E6 F: FTMS + p ESI Full ms [120.00-1... 295.1056 $C_{17}H_{15}O_3N_2$ = 295.1077 11.5 RDBE -7.2090 ppm 296.1085	**12**
0.81	CPS_HCOOH_FT_170323145140 #752-755 RT: 6.32-6.32 AV: 2 SB: 2 6.25 , F: FTMS + p ESI Full ms [100.00-1... 344.0794 $C_{17}H_{15}O_3N_3Cl$ = 344.0796 11.5 RDBE -0.6562 ppm 158.9612 238.2019	**13**

RRT		
1.14		
1.69		
1.93		

Table 3.3. Main impurities present in the product obtained by coupling. RRT = relative retention time (ratio between the retention time of the impurity and retention time of Lenvatinib).

The structure of all the impurities was hypothesized from MS spectra reported in Table 3.3. In the case of impurities **10-13, 15** and **16** the main peaks corresponds to [M+H]⁺ or [M+Na]⁺ or [M+K]⁺. The MS spectrum of impurity **14** shows two peaks: m/z = 357 which corresponds to the double charged ion [M+H]²⁺ and m/z = 735 which corresponds to [M+Na]⁺. Furthermore, the impurities **10-14** were synthesized and characterized (NMR, MS, FTIR) confirming the hypothesized structure (Paragraph 3.5). Therefore, for these impurities the limit in the final API is 0.15% according to ICH guidelines.

The impurities **10** and **11** are favored by the presence of water traces in the coupling step, but working under a nitrogen atmosphere they are usually kept below 0.10%. Thus, they could be considered not critical at this stage, while impurity **11** has a key role during the formation of the mesylate salt, as it will be discussed in paragraph 3.3.

The impurity **12** can be originated by reaction of 4-chloro-7-methoxy-quinoline-6-carboxamide and phenol in the coupling step. As described in the previous paragraph, phenol is an impurity of urea **3** and it's usually present below 0.2% (according to HPLC analysis of urea **3** at 271 nm). Moreover, it's soluble in several solvents and hence it does not represent a critical impurity.

56

The structures **13**, **14** and **15** are generated in the coupling step probably due to degradation of Lenvatinib in base at the reaction temperature. Indeed, heating a DMSO solution of Lenvatinib and Cs_2CO_3 at 70 °C for 72 h determines a significant increase of these impurities (Table 3.4).

Compound	A % Starting Lenvatinib	A % After 72 h
Impurity **13**	0.20	14.41
Lenvatinib	98.04	66.86
Impurity **14**	0.45	8.94
Impurity **15**	0.36	1.35

Table 3.4. Degradation of Lenvatinib DMSO, Cs_2CO_3 at 70°C. Values determined by HPLC A %.

Based on this observation, it is possible to hypothesize a plausible mechanism for the formation of the impurities in the coupling step (Scheme 3.5).
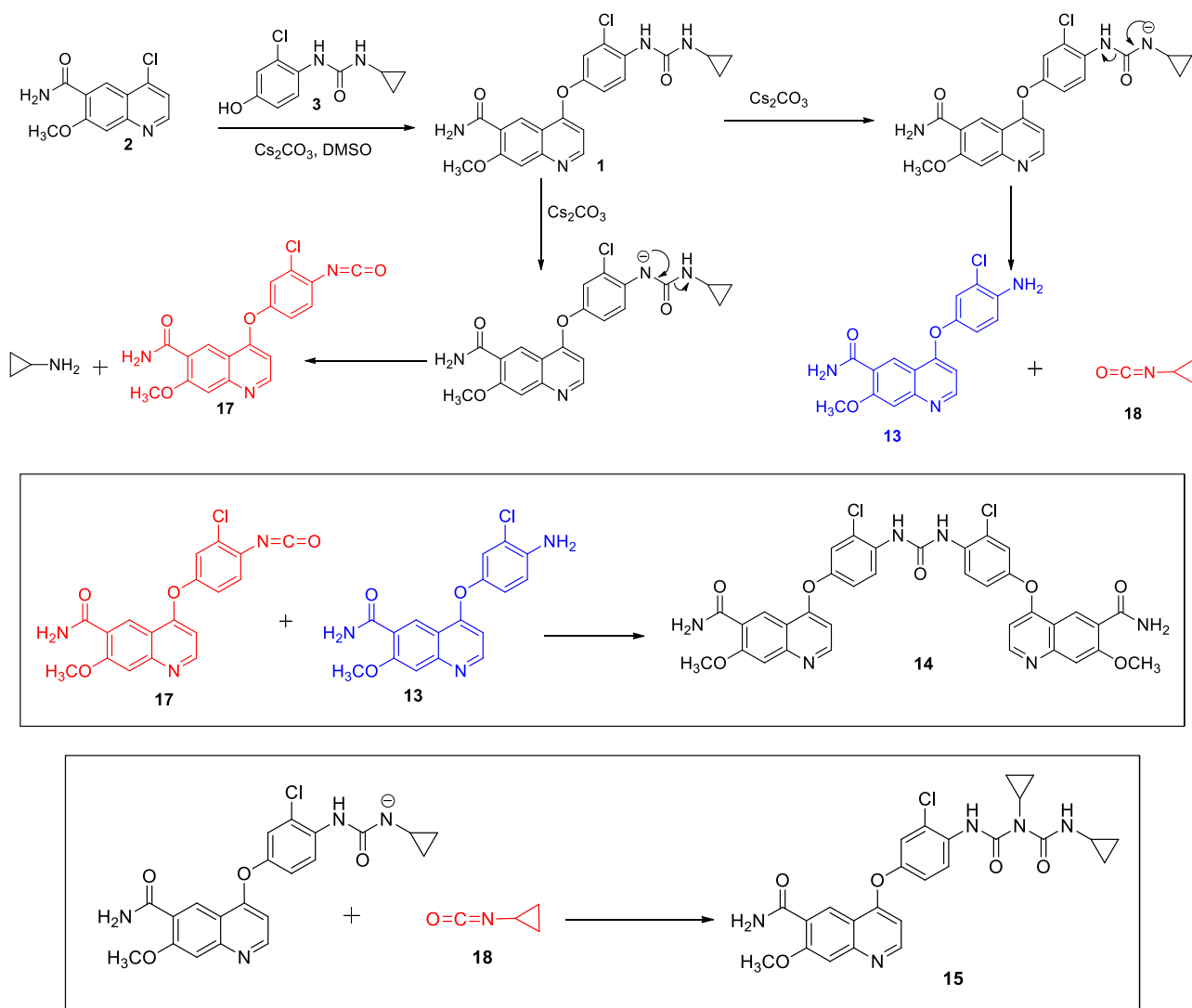
Scheme 3.5. Plausible mechanism of formation of impurities 13, 14 and 15.

The urea moiety in Lenvatinib **1** can be deprotonated by cesium carbonate. When the deprotonation occurs to the nitrogen attached to the phenyl ring, it induces the elimination of cyclopropylamine and the formation of an aromatic isocyanate **17** (red figure). On the other hand, when the nitrogen attached to cyclopropylamine is deprotonated, cylcopropylisocyanate **18** (red figure) results and impurity **13** is formed. If isocyanate **17** reacts with amine **13**, the dimer **14** is obtained. Moreover, the formation of impurity **15** can be explained by reaction of Lenvatinib deprotonated in the aliphatic nitrogen position with cylopropylisocyanate **18**. To minimize the formation of these three impurities a further optimization of coupling conditions was required and will be discussed in paragraph 3.2d.

Finally, impurity **16** is characterized by the presence of a nitrile group instead of an amide in position 6 of the quinoline ring and it is related to the quality of the starting material. Indeed, 4 chloro-quinoline **2** can be contaminated by nitrile **19**, which is transformed into impurity **16** in the coupling step (Scheme 3.6).
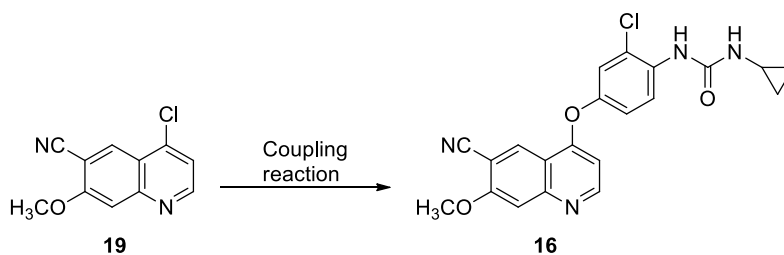
Scheme 3.6. Plausible origin of impurity 16.

Impurity **16** is very difficult to remove by crystallization of Lenvatinib **1** and hence to have this impurity below 0.10% in the final product it is necessary to control the content of 4-chloro-7-methoxyquinoline-6-carboxamide **19** in starting quinoline **2** (paragraph 3.2c).

3.2c Quinoline purification

4-chloro-7-methoxyquinoline-6-carboxamide **2** is commercially available and two different batches of two different producers were purchased in our lab. The batch produced from Tianjin Pharmaceutical has HPLC purity of 97.73 % at 309 nm and HPLC purity of 98.97% at 250 nm. At the second wavelength a peak at RRT = 2.20 with A% = 0.23 was clearly present. According to MS and MS-MS spectrum this peak can be identified as 4-chloro-7-methoxyquinoline-6-carbonitrile **19,** which is probably an intermediate of the synthesis of quinoline **2.** Indeed, according to the synthetic route described in the originator's patent[1], the amide in position 6 of quinoline ring is

introduced starting from a nitrile in that position, which is first hydrolysed to carboxylic acid and then converted to amide via acyl chloride (Scheme 3.7).
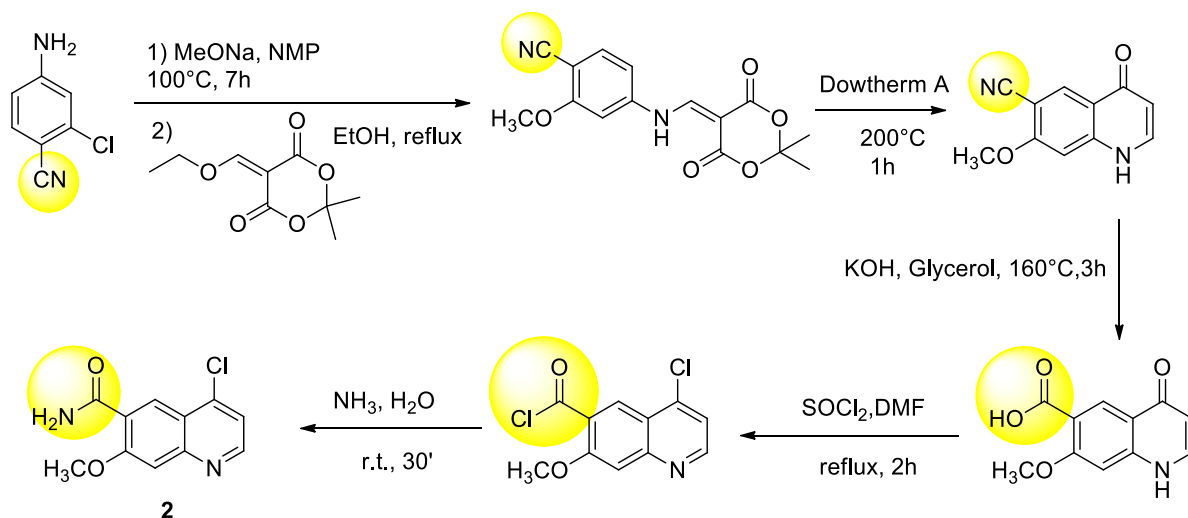
Scheme 3.7. Synthesis of starting 4-Chloro-quinoline 2.

On the other hand, the second batch purchased from Shandong Boyuan Pharmaceutical has a higher HPLC purity and impurity **19** is completely absent probably because the amide in position 6 of the quinoline ring is introduced with a different strategy.

Anyway, to have a robust process from any possible starting material, we decided to introduce a purification step of the quinolone, able to remove the impurity **19**.

Several trials of recrystallization were performed on 2-3 g of quinoline **2** purchased from Tianjin Pharmaceuticals and the solids were analysed by HPLC at 250 nm (Table 3.5).

Exp.	Description	Quinoline	RRT = 2.13	Impurity 19	RRT =2.24	Recovery (%)
0	Quinoline from Tianjin	98.97	0.12	0.23	0.34	-
1	9:1 DMSO:H$_2$O	99.33	0.13	0.13	0.32	80
2	1:2 DMSO:CH$_2$Cl$_2$	99.53	0.12	0.12	0.13	87
3	BuOH	98.58	0.16	0.25	0.37	78
4	2-MeTHF	98.92	0.13	0.22	0.33	89
5	DMC	98.90	0.12	0.23	0.33	89
6	CH$_3$CN	98.81	0.13	0.22	0.39	91
7	NMP	99.61	0.07	0.06	0.18	84

Table 3.5. Experiments of quinoline purification. The values are HPLC A% (250 nm) of solids after recrystallization. RRT = relative retention time.

4-chloro-7-methoxyquinoline-6-carboxamide **2** is soluble in 6V of DMSO at 100°C, so two trials were performed by dissolution in DMSO followed by addition of antisolvents (Water, dichloromethane). In these two conditions (Exp. 1 and Exp. 2) the content of impurity **19** was reduced from 0.23 to 0.12-0.13 %. Other experiments were performed in Butanol (BuOH, Exp. 3), 2-methyltetrahydrofuran (2-MeTHF, Exp. 4), dimethylcarbonate (DMC, Exp. 5) and acetonitrile (CH₃CN, Exp. 6). In these solvents quinoline **2** was not completely soluble even at high temperature, but the suspension was heated at reflux for 1h and then cooled to r.t. However, no increase in purity of quinoline **5** was observed in all these trials. In Experiment 6 quinoline **2** was dissolved in *N*-methyl-2-pyrrolidone (NMP) at 100°C and then was cooled to room temperature for 18h and at 0°C for 1h to crystallize the solid without addition of antisolvent. Impurity **19** was reduced from 0.23 % to 0.06 % and the recovery was satisfactory (84%). Therefore, we used the recrystallization in *N*-methyl-2-pyrrolidone to standardize the quality of starting 4-chloro-7-methoxyquinoline-6-carboxamide **2.** Other effective alternatives can also be the strategies based on dimethyl sulfoxide and anti-solvents.

3.2d Optimization of coupling conditions

As discussed in paragraph 3.2b, the impurities **13**, **14** and **15** were formed during the coupling reaction and hence optimization of coupling conditions was necessary to minimize them, especially impurities **14** and **15**, which are hard to remove by recrystallization of Lenvatinib. On the other hand, impurity **13** and the unreacted quinoline **2** are easier to remove in the following steps. First of all, we tried to change the solvent of the reaction from dimethylsulfoxide (Entry 1) to dimethylformamide (Entry 2) in order to make the product less soluble and hence to disfavor its degradation (Table 3.6). The results were evaluated determining the impurities content by HPLC.

Entry	Solvent	T (°C)	t (h)	% Lenv. 1	% Imp 13	% Imp 14	% Imp 15	% quinoline 2
1	DMSO	70	24	95.00	1.81	0.46	1.74	0.22
2	DMF	70	24	98.43	0.04	0.19	0.06	0.40

Table 3.6. Change of solvent in coupling step. All these experiments were performed with 2-3 g of quinoline **2** and 1.2 eq. of urea **3**.

The use of dimethylmorfamide does not affect the reaction rate and allows to decrease the impurities significantly. However, when we tried to scale-up the reaction in these conditions from

2 g to 15 g, the improvement was not confirmed and the impurities content was comparable to the dimethylsulfoxide experiment.

A set of trials was then performed to evaluate the effect of reducing the amount of Cs_2CO_3, which could play a key role in the formation of the impurities (Table 3.7).

Entry	Solvent	T (°C)	t (h)	Cs_2CO_3 eq.	% Lenv. 1	% Imp. 13	% Imp 14	% Imp 15	% quinoline 2
1	DMSO	70	24	2	95.00	1.81	0.46	1.74	0.22
2	DMSO	70	24	1.2	94.70	1.73	1.17	0.08	0.22
3	DMSO	70	24	0.9	95.90	0.22	0.32	0.10	2.24
4	DMSO	70	48	0.6	95.15	0.51	0.19	0.06	2.51

Table 3.7. Effect of reduction of Cs_2CO_3 equivalent in coupling. All these experiments were performed with 2-3 g of quinoline **2** and 1.2 eq. of urea **3**.

The first experiment (Entry 1) was performed in the conditions reported in the originator's patent and the percentages of impurities **14** and **15** were very high (0.46 and 1.74 % respectively). Reducing Cs_2CO_3 from 2 to 1.2 mol. equiv. (i.e. the same amount as urea **3**) (Entry 2), decreased the content of impurity **15** significantly, but did not have a good effect on the other impurities. In experiments 3 and 4 cesium carbonate was reduced to 0.9 and 0.6 equivalents, respectively. In these cases, we wanted to achieve deprotonation of a portion of phenol groups of the urea **3** by bicarbonate generated from the reacted cesium carbonate. The results show that the reaction proceeds much more slowly (more than 2% of unreacted quinoline was present at the end of the reaction in both the experiments) and the formation of impurity **14** was not sufficiently limited.

Another set of experiments were run to study the effect of temperature on the reaction course and purity of product by HPLC A% (Table 3.8).

Entry	Solvent	T (°C)	t (h)	Cs_2CO_3 eq.	% Lenv. 1	% Imp 13	% Imp 14	% Imp 15	% quinoline 2
1	DMSO	70	24	1.2	94.70	1.73	1.17	0.08	0.22
2	DMSO	60	24	1.2	97.31	0.43	0.12	0.12	2.51
3	DMSO	50	24	1.2	89.09	0.05	0.11	0.12	9.20

Table 3.8. Effect of the reduction of reaction temperature. All these experiments were performed with 2-3 g of quinoline **2** and 1.2 eq. of urea **3**.

These trials show that the reduction of the reaction temperature has a significant effect on the formation of impurity **14**, which is reduced from 1.17% at 70°C (Entry 1) to 0.11% at 50°C (Entry 3).

Moreover, impurity **15** is maintained at low levels in all the experiments probably thanks to the use of 1.2 equivalents of Cs_2CO_3. However, at 50 °C the reaction becomes too slow and after 24 h, 9.20 % of the starting quinoline **2** is still present in the precipitate at the end of the reaction.

From the last two sets of experiments (Table 3.7 and 3.8) we can observe that a stoichiometric or sub-stoichiometric amount of cesium carbonate is required to reduce impurity **15**, while the temperature is the critical parameter to control the formation of impurity **14.** On the other hand, we need conditions that do not slow down the reaction rate too much. Therefore, we designed an experiment with 1 equivalent of quinoline **2**, 2 equivalents of urea **3**, 2 equivalents of Cs_2CO_3 (stoichiometric amount relative to urea) dissolved in 6 V of DMSO and heated at 50 °C for 24 h. The HPLC profile of the precipitate recovered from the reaction in these conditions is summarized in Table 3.9 and compared to the precipitate obtained according to the originator's conditions.

Impurity Identification	Originator process	Indena process
Impurity **13**	1.81	0.16
Quinoline **2**	0.22	0.95
Lenvatinib	95.00	97.39
Impurity **14**	0.46	0.07
Urea **3**	0.04	0.08
Impurity **15**	1.74	0.12

Table 3.9. Comparison of HPLC profile of coupling precipitate obtained with originator and Indena conditions.

With the conditions described above the formation of impurities was significantly minimized: impurity **13** was reduced from 1.81% to 0.16%; impurity **14** was decreased from 0.46% to 0.07 and impurity **15** was limited from 1.74% to 0.12%.

3.2e Recrystallization of Lenvatinib free base

The precipitate recovered from this coupling reaction must be recrystallized in order to decrease the amount of unreacted quinoline **2** and impurity **15** and hence to satisfy the purity requirements for our generic API. We evaluated the solubility of Lenvatinib free base **1,** which was soluble only in dimethylsulfoxide (DMSO), dimethylformamide (DMF) and Acetic acid (AcOH). We tested different recrystallization methods on few milligrams of crude product by dissolving Lenvatinib **1** in these three solvents at high temperature and by adding an anti-solvent (water, dichloromethane or ethyl acetate). Among all the solvent mixtures the most promising ones seemed to be 4:1

DMSO : water and 1 : 3 DMSO : CH$_2$Cl$_2$ and thus we tested them on a few grams of crude Lenvatinib evaluating the purity of the product obtained after recrystallization (Table 3.10).

Exp.	0	1	2
Impurities	Coupling precipitate	Recrystallization in 4:1 DMSO : water (y=76%)	Recrystallization in 1:3 DMSO:DCM (y= 73%)
Impurity **12**	0.18	0.03	/
Impurity **13**	0.16	/	/
Quinoline **2**	0.95	0.26	0.20
Lenvatinib 1	97.39	99.11	99.65
Impurity **14**	0.07	0.07	0.05
Impurity **15**	0.12	0.12	0.05
Impurity **16**	0.09	0.08	0.06

Table 3.10. Comparison between HPLC profiles of Lenvatinib free base after two different recrystallization methods.

According to the data in Table 3.10, impurities **12** and **13** are effectively removed by both the methods of recrystallization. On the other hand, the use of dichloromethane as anti-solvent (Exp. 2) allows a slightly higher reduction of the unreacted quinoline, which passes from 0.95% to 0.20%. Moreover, this method reduces the critical impurity **15** from 0.12 % to 0.05% and increases the percentage of area of Lenvatinib **1** from 97.39% to 99.65%, a value higher than the purity achieved with the water method (Exp. 1).

Therefore, we decided to recrystallize Lenvatinib free base **1** with 1:3 DMSO:DCM, because it allows an higher purification especially from the critical impurities. At this stage the product has all the impurities below ICH limit, but quinoline **2**, which is 0.20 % and will be removed in the last step (the crystallization of the mesylate salt).

3.3 Preparation of mesylate salt of Lenvatinib (ACA-1-HT-dry)

The last step in the preparation of the API is the formation of the mesylate salt of Lenvatinib (Scheme 3.8). Moreover, the mesylate salt has to be crystallized to afford ACA-1-HT dry solid form, which was deeply discussed in detail in Chapter 2 and is essential to be not infringing to solid form patent.

The mesylate salt was formed by suspending Lenvatinib free base **1** in 2.7 V of AcOH and adding drop-wise 1 eq. of methanesulfonic acid dissolved in 0.3 V of acetic acid. The mixture was heated at 60 °C until complete dissolution and 1 V of Ethyl acetate was added. Then, the solution was

cooled to 40°C. The cooling rate is a critical parameter to obtain the desired polymorph. Indeed, a fast cooling determines the crystallization of product with originator's Form I, while a temperature decrease from 60°C to 40°C in about 1.5 h allows to obtain ACA-1 form. Finally, the acetic acid solvate ACA-1 is converted into anhydrous ACA-1-HT-dry by drying.
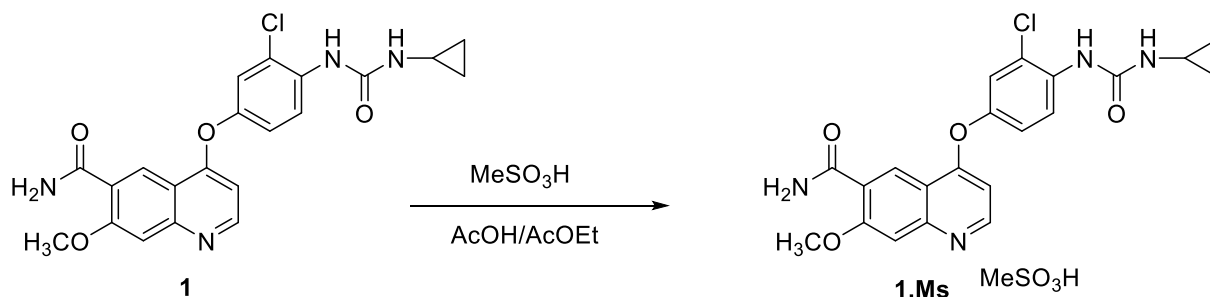
Scheme 3.8. Preparation of mesylate salt of Lenvatinib (ACA-1-HT-dry).

The drying process is also necessary to remove the solvents below the limits described in ICH guidelines for residual solvents.[6] Acetic acid is the most difficult one to remove because of its high boiling point and drying at 70°C for 5 days is necessary to reduce its content in final API below 5000 ppm.

As described above, the formation of the mesylate salt requires acidic conditions, which can promote the degradation of Lenvatinib to impurity **11** (Table 3.3). Degradation of Lenvatinib occurs mainly during the addition of methanesulfonic acid. Indeed, the suspension of Lenvatinib free base in acetic acid is not homogenous and easy to stir and consequently methanesulfonic acid is not well dispersed. The local presence of excess of MeSO$_3$H can determine the degradation of Lenvatinib to impurity **11**. To solve this issue, we din not charge in the reactor all the acetic acid, but only 1/2 of the total amount of Lenvatinib **1** and then 1/2 of MeSO$_3$H was added drop-wise. The suspension was stirred at r.t. for 15' to dissolve the major part of the solid before adding the second half of Lenvatinib **1** and MeSO$_3$H. Splitting the addition of the reagents allows to have always an homogenous suspension in the reactor and hence to avoid the local accumulation of acidity. With this method it was possible to minimize the formation of impurity **11** significantly below 0.15%, the ICH limit for a characterized impurity.

3.4 Process development and scale-up

In the previous paragraphs the optimization of process conditions (reaction parameters and methods of purification) was presented and the main aim of the work was to achieve the API purity required by regulatory entities. A typical carry-over of the impurities with the optimized process is shown in Table 3.11.

| Impurities			Precipitate ater coupling (1:1 DMSO : H₂O)	Free base recrystalized in 1:3 DMSO : DCM	Lenvatinib Mesylate in AcOH : AcOEt	Specification limit
n	**RRT**	**MW**				
10	0.45	218	/	/	0.04	0.15
11	0.61	386	0.03	/	0.05	0.15
12	0.80	294	0.10	/	0.04	0.15
13	0.81	343	0.10	0.04	0.02	0.15
2	0.92	236	1.28	0.21	/	0.15
1	**1.00**	**426**	**97.34**	**99.60**	**99.72**	-
14	1.14	712	0.11	0.05	0.04	0.15
-	1.16	/	0.09	/	/	0.10
3	1.23	226	0.03	/	/	0.15
-	1.27	427	0.10	/	/	0.10
-	1.31	/	0.05	/	/	0.10
15	1.69	509	0.18	0.05	0.03	0.10
16	1.93	408	0.06	0.05	0.02	0.10
-	2.22	/	0.04	/	/	0.10

Table 3.11. Impurities carry-over with optimized process. Data are calculated as HPLC Area %.

In the last column of Table 3.11 there are the specification limits of the impurities according to ICH guidelines: the unidentified impurities must be below 0.10%, while the identified and characterized impurities (compounds **10, 11, 12, 13** and **14** and the intermediates **2** and **3**) must be below 0.15%. The new coupling conditions allow to keep under control the formation of the most critical impurities and the recrystallization method increases further the product purity and decreases the content of the unreacted quinoline **2** and of the impurities **14** and **15.** In Lenvatinib mesylate obtained with the optimized process all the impurities are below the limits and hence the product satisfy the quality requirements for a generic API.

Before the scale-up and industrialization of the process two additional practical issues must be tackled: the filtration of the precipitate at the end of the coupling step and the discharge of the final suspension of Lenvatinib mesylate from the reactor. Indeed, at the end of the coupling reaction the DMSO mixture is cooled to room temperature and the product is precipitated by addition of water. However, the suspension is not fluid and hard to stir and the following filtration is pretty slow. Moreover, we observed that at the end of the reaction in dimethylsulfoxide there is a precipitate which might be cesium salts generated from Cs_2CO_3 ($CsHCO_3$ and/or $CsCl$). Therefore, the precipitated inorganic salts were removed by filtration and the filtrate was further diluted with DMSO before the precipitation with water. The removal of inorganic salts and the dilution of the

reaction mixture allowed to obtain a suspension much more fluid which can be filtered more easily.

The second issue is related to the last step of crystallization of Lenvatinib mesylate in 1:3 AcOH : EtOAc. At the end of the crystallization the final suspension is characterized by high density and is similar to a wet solid and hence it can not flow out from the reactor. To solve this problem, we tried to perform the crystallization in a more dilute suspension (6 V of AcOH instead of 3 V and 2 V of EtOAc instead of 1V), but the product was obtained in Form I. The same was achieved by maintaining constant the amount of acetic acid (3V) and using more ethyl acetate (3 V). Therefore, we decided to carry out the crystallization in the classical conditions (3 V of AcOH e 1 V of EtOAc) and to add other 5 V of ethyl acetate once the solid was precipitated. The addition of fresh ethyl acetate allowed to make the suspension more fluid and to flow it out from the reactor. The stability of ACA-1 form in these conditions was controlled by XRPD (Figure 3.2).
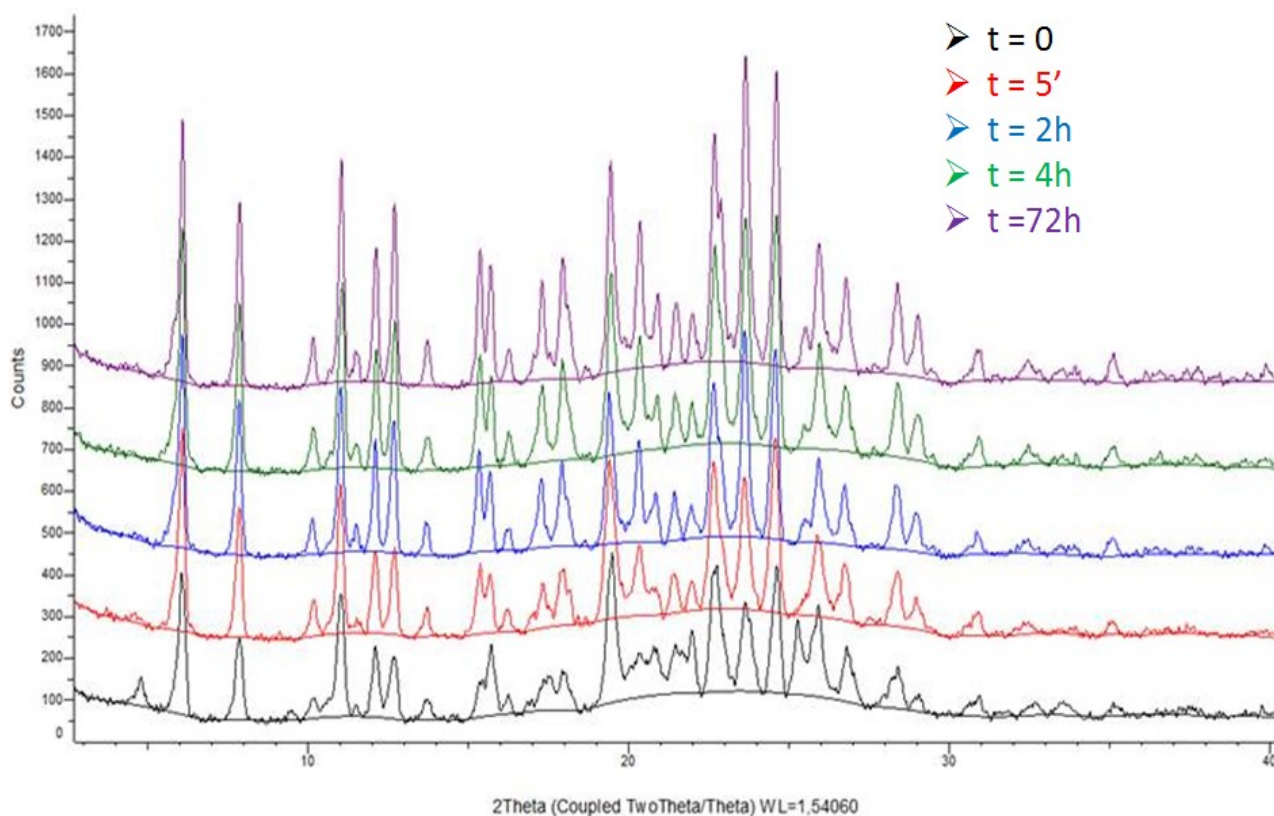
Figure 3.2. XRPD overlap of solid from slurry in AcOH (3V) and EtOAc (6V) at different time.

As we can see in Figure 3.1, the diffractogram remains almost unchanged in time and the only differences are in the peaks at 4.5° and 9.5° 2-theta, which are present in the black profile (t = 0) and are absent in all the other ones. They can be attributed to traces of Form I (originator's acetic acid solvate), which probably converts to ACA-1 by slurry in AcOH (3V) and EtOAc (6V). Therefore,

the addition of ethyl acetate at the end of the crystallization is a good technique to recover the suspension of Lenvatinib-ACA-1 from the reactor without any conversion to patented solid forms.

Finally, the optimized process was scaled up in a high containment Kilo-Lab equipped with 25 L vessels, filter and oven in glove boxes (Figure 3.3). I n this plant 1 Kg of Lenvatinib mesylate-ACA-1-HT-dry form was prepared from 4-amino-3-chlorophenol and 4-chloro-7-methoxyquinoline-6-carboxamide. The product was obtained with 99.70 % of HPLC purity and all the impurities below 0.10%.

Figure 3.3. High containment Kilo-Lab (25 L vessel and oven).

3.5 Synthesis and characterization of impurities

As anticipated before, in collaboration with Università degli studi di Camerino (Prof. Palmieri) we synthesized and characterized the critical process-related impurities (compounds **10, 11, 12, 13** and **14**).

Impurity **10** is characterized by a hydroxyl group attached in position 4 of quinoline ring. It was synthesized from benzyl alcohol **20** which was deprotonated by sodium hydride and then coupled to 4-chloro-7-methoxyquinoline-6-carboxamide **2** to afford intermediate **21** (Scheme 3.9). In the second step the benzyl group was removed by reduction with ammonium formate and 10% Pd/C.

At the end of the reaction the catalyst was removed and the crude product **10** was purified by flash chromatography to achieve 98.8% of HPLC purity.
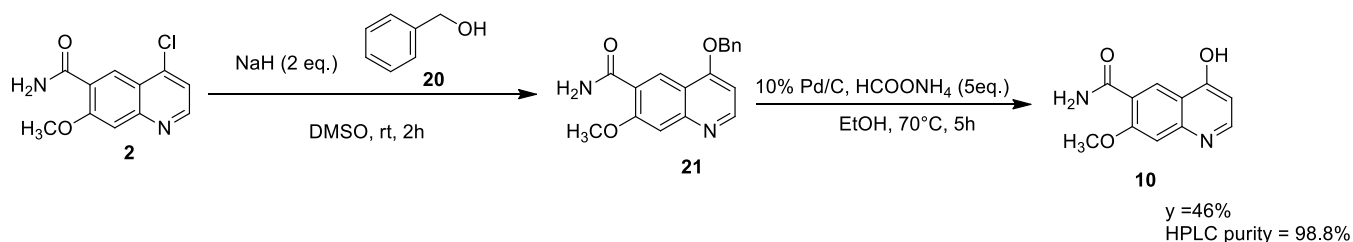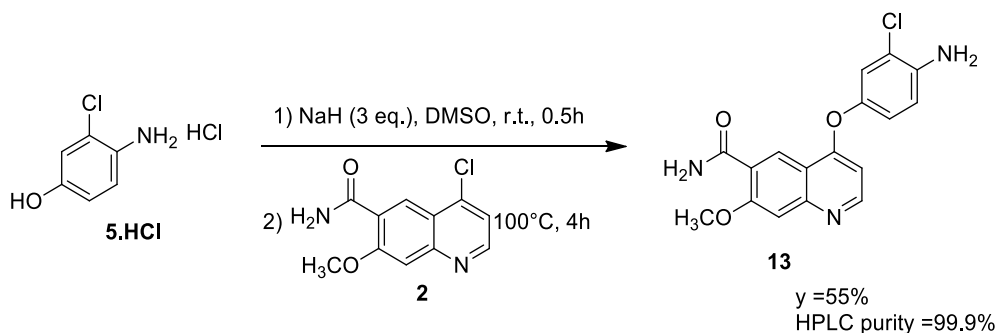

Scheme 3.9. Synthesis of impurity 10.

Impurity **12** has a phenoxy group attached in position 4 of quinolone ring and can be synthesized with a strategy similar to the preparation process of Lenvatinib (Scheme 3.10). Indeed, phenol **22** is coupled to 4-chloro-quinoline **2** in DMSO and in the presence of Cs_2CO_3 as base. At the end of the reaction the product **12** is isolated by precipitation of water with HPLC purity of 99.0% and hence an additional purification step is not necessary.
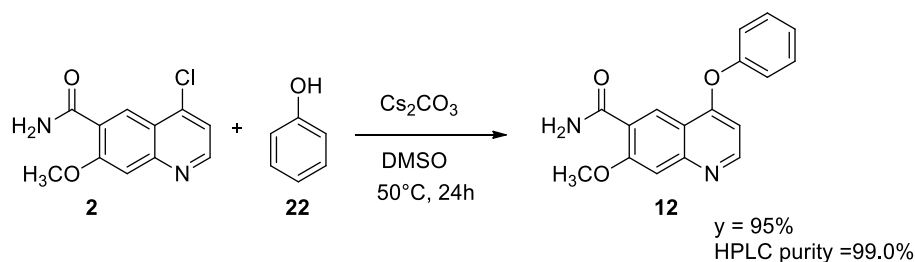
Scheme 3.10. Synthesis of impurity 12.

Impurity **13** is characterized by the absence of cyclopropylurea moiety and can be synthesized by coupling of 4-chloro-quinoline **2** with 4-amino-3-chloro-phenol hydrochloride salt **5.HCl** (Scheme 3.11). The reaction was carried out in DMSO at 100°C and after 4h the mixture was cooled to room temperature and diluted with water. The product **13** was extracted with ethyl acetate and purified by flash chromatography to achieve 99.9 % of HPLC purity.

Scheme 3.11. Synthesis of impurity 13.

68

The impurity **13** can also be exploited as starting material for the preparation of impurities **11**, which has a primary urea group without cyclopropyl, and **14**, which has a dimeric structure (Scheme 3.12).
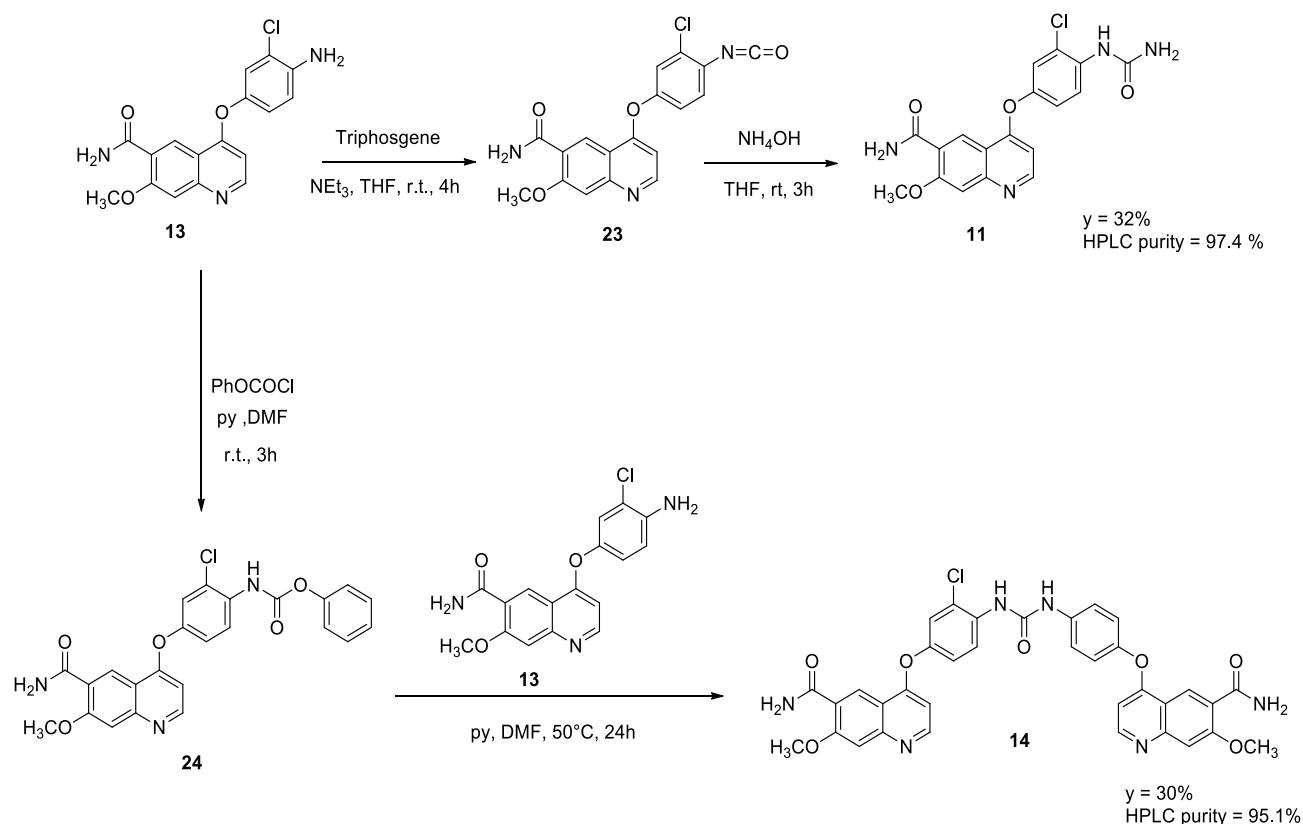
Scheme 3.12. Synthesis of impurity 11 and 14.

On one side, compound **13** reacted with triphosgene to afford isocyanate **23**, which was then converted into the corresponding urea with ammonium hydroxide. Both the steps were monitored by FTIR focusing on the isocyanate band at 2265 cm^{-1} and the final product **11** was purified by column chromatography to obtain 97.4% of HPLC purity. On the other side, the amine **13** was transformed into phenyl carbamate **24** according to a procedure reported in originator's patent US 7683172 (Phenyl chloroformate, pyridine as base and DMF as solvent).[1] Then, the intermediate **23** reacted with a second equivalent of amine **13** to yield the dimer **14**, which was purified by chromatography (HPLC purity = 95.1%).

The synthesis of all the impurities, but compound **12**, were carried out on a few milligrams and the yields were modest (y = 30-50%). None of the steps described before was optimized because our goal was only to obtain a small amount of sample to clearly identify and characterize the impurities present in our API. The identification and characterization were achieved by ^{1}H-NMR, ^{13}C-NMR, ESI-MS, FTIR and HPLC and all the data are reported in the

experimental section. Thanks to this work, the limit of impurities **10, 11, 12, 13** and **14** in the final API was raised to 0.15% in accordance with ICH guideline.[3]

Experimental section

4-Amino-3-chloro-phenol hydrochloride salt and 4-chloro-7-methoxyquinoline-6-carboxamide were purchased from Tianjin Pharmaceutical, Phenylchloroformate from Xhandong Tianan Chemical and cyclopropylamine from Zhejiang Shaxing Pharmaceutical. 2-Methyltetrahydrofuran (2-MeTHF), N-methyl-2-pyrrolidinone (NMP), methanesulfonic acid and acetic acid were purchased from Sigma Aldrich – Merck, while the other solvents (EtOAc, DCM) were technical grade solvents commonly used in Indena plants.

Nuclear magnetic resonance spectra were obtained on Agilent MR400 DD2 VNMRJ™ Software Rev. 4.2. All NMR experiments were recorded at 400 MHz for ^{1}H and 100 MHz for ^{13}C with an OneNMR probe (1H 90° pulse width, 13C 45° pulse width) at a temperature of 30 °C.

About 10 mg of sample were dissolved in 0.75 ml of solvent (into a 507-PP tube 7 inches x 5 mm. Chemical shifts are referred to tetramethylsylane (99.7% UVASOL Merck, code 8183) at 0 ppm as external standard.

Spectra were acquired using 49 K data points and spectral width of 16 ppm and 240 ppm for ^{1}H and ^{13}C respectively, 65 K data points were used for the processing with lb=2 for ^{13}C.

Mass spectroscopy was carried out with LTQ-Orbitrap Mass Spectrometer equipped with an ESI interface and Excalibur Data System (Indena code LTQ-Orbitrap) coupled with Infinity 1290 UHPLC Agilent. MS Analysis was carried out in full scan mode operation at a resolution of 30.000, recording MS spectra both in negative and positive mode.

MS parameters were set as follows:

	ESI positive
Capillary temperature	450°C
Sheath gas flow	23 au
Auxiliary gas flow	2 au
Source voltage	+4.5 KV
Capillary voltage	+17.50 V
Tube Lens	+109 V

The infrared spectrum was recorded in Attenuated Total Reflectance (ATR) mode using Fourier-Transform spectrometer Nicolet iS10 Thermo Scientific, equipped with Specac ATR Golden Gate and OMNIC software (Thermo Scientific). The spectrum was the result of the acquisition and transformation of 32 co-added scans in the 4000-550 cm^{-1} spectral region at a resolution of 4 cm$^{-1.}$

HPLC system consists of Quaternary pump, thermostated cooling autosampler, thermostated column compartment and UV/VIS detector connected to Empower software.

Stationary phase: Phenyl-Hexyl.

Mobile phase: 0.012% TFA in Water and 0.012% TFA in acetonitrile (linear gradient).

Other details of HPLC methods (e.g. flow rate, detection, injection volume, Column temperature) are not reported here for trade secret reasons.

Residual solvents were determined by Head space Gas chromatography.

Water content was determined by Karl Fischer- Coulometric Titration, according to Ph. Eur. 2.5.32. Test material is heated in an oven at suitable temperature, water is released from the solid matrix, evaporated and carried into the cell by a stream of inert dry gas (evaporation oven sample technique).

Synthesis of 1-(2-chloro-4-hydroxyphenyl)-3-cyclopropylurea (*3*)
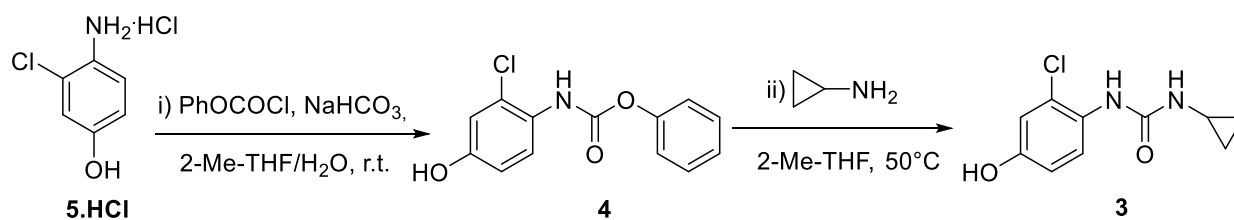
4-Amino-3-chloro-phenol hydrochloride salt **5.HCl** (60.0g, 333.3 mmol, 1 eq.) was suspended in 2-methyltetrahydrofuran (180 mL, 3V) and the suspension cooled at 0÷5°C. A solution of NaHCO$_3$ (58.8 g, 699.9 mmol, 2.1 eq.) in water (650 mL) was added drop-wise in 25 min. and the temperature maintained below 10°C. A solution of phenyl-chloroformate (57.4 g, 46.0 mL, 366.7 mmol, 1.1 eq.) in 2-methyltetrahydrofuran (96 mL) was added drop-wise in 25 min. keeping the temperature below 10°C. The mixture was stirred at 0÷5°C for 10 min. The reaction was monitored by quantitative TLC (95:5 DCM : MeOH; UV 254 nm) until 4-amino-3-choloro-phenol was less than 1%.

The phases were separated and removed and cyclopropylamine (37.9 g, 46.0 mL, 666.6 mmol, 2 eq.) was added drop-wise in 30 min to the organic phase maintaining the temperature below 10°C. The mixture was stirred at 50°C for 3h. The reaction course was monitored by quantitative TLC (7: 3 Hexane : EtOAc, UV 254 nm) until phenyl carbamate intermediate **4** was less than 0.5%.

The reaction mixture was cooled to r.t. and 1M H$_2$SO$_4$ (200 mL) was added in 15 min. The aqueous phase was removed and the organic phase diluted with 2-methyltetrahydrofuran (60 mL, 1V) and washed with 10% NaCl (200 mL). The organic phase was concentrated to 120 mL (2V) at T$_{ext}$ = 50 °C and P= 150 mbar. Ethyl acetate (240 mL, 4V) was added and the resulting mixture concentrated to 240 mL (4V). Ethyl acetate (120 mL, 2V) was added and the mixture concentrated to 240 mL (4V). The latter operation was repeated twice. The obtained slurry was stirred at room temperature for 1h and *n*-heptane (60 mL, 1V) was added. After 30 min., the precipitate was filtered, washed with 4:1 EtOAc: *n*-heptane (60 mL, 1V) and with *n*-heptane (60 mL, 1V). The white solid was dried at 55°C under vacuum for 16h to yield 1-(2-chloro-4-hydroxyphenyl)-3-cyclopropylurea **3** (68.5 g, y = 91%).
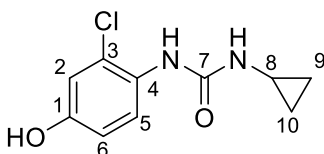
<u>1H-NMR</u> (DMSO-d$_6$) δ (ppm): 0.38 (2H, m, H$_{9A}$,H$_{10A}$), 0.61 (2H, m, H$_{9B}$,H$_{10B}$), 2.50 (1H, m, H$_8$), 6.65 (1H, dd, *J* = 8.9 Hz, *J* = 2.8 Hz, H$_6$), 6.77 (1H, d, *J* = 2.8 Hz, H$_2$), 6.79 (1H, d, *J* =2.6 Hz, NH$_8$), 7.53 (1H, s, NH$_4$), 7.68 (1H, d, J = 8.9 Hz, H$_5$), 9.44 (1H, s, OH).
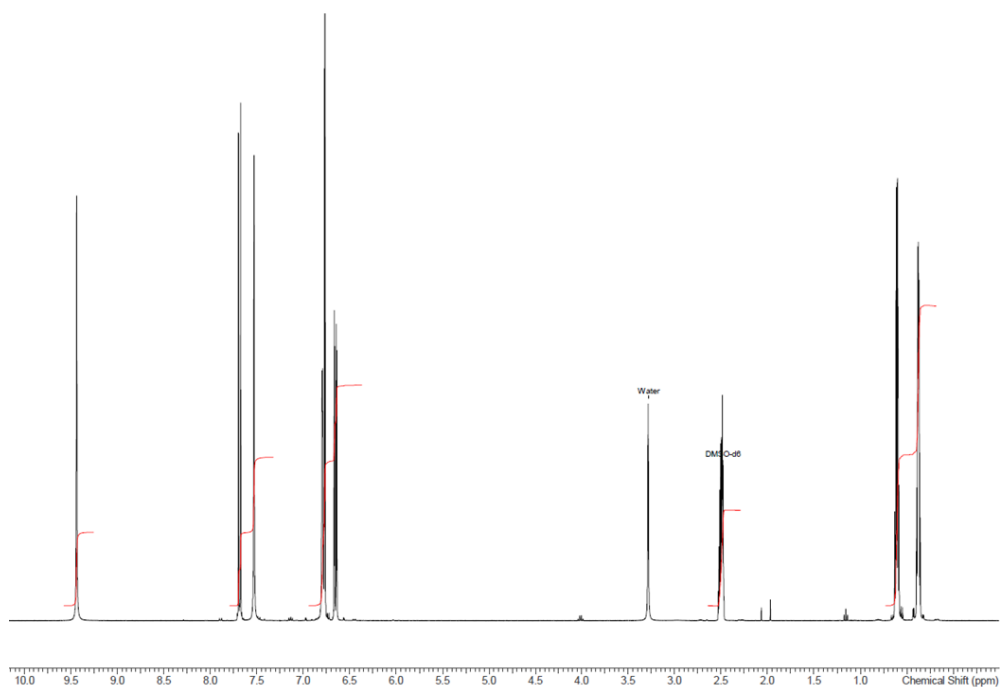
<u>13C-NMR</u> (DMSO-d$_6$) δ (ppm): 6.8 (C$_9$, C$_{10}$), 22.8 (C$_8$), 114.8 (C$_2$, C$_6$), 115.7 (C$_5$), 124.4 (C$_3$), 128.5 (C$_4$), 153.5 (C$_1$), 156.4 (C$_7$).
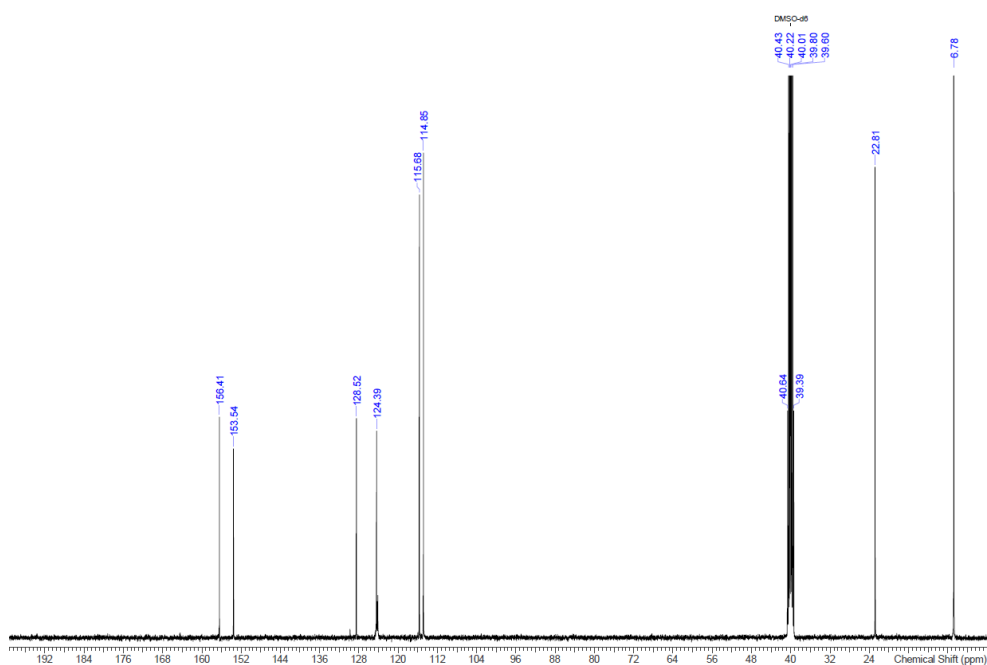
<u>ESI-MS</u>: m/z= 227.1 [M+H]$^+$

HPLC chromatogram (λ = 285 nm): Typical HPLC purity is 99.88 %.
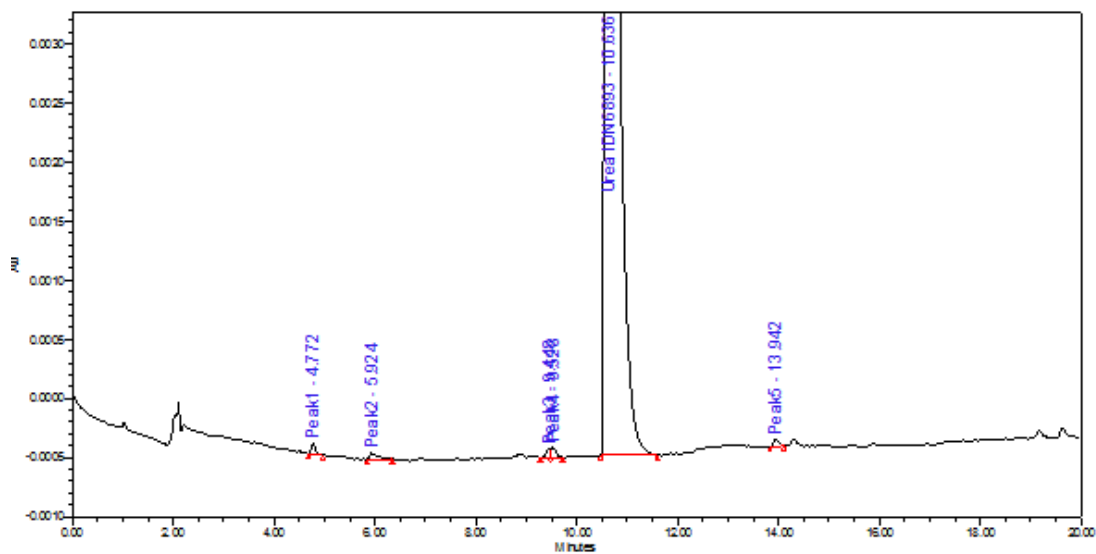
	Name	Int Type	RT (min.)	RRT	Area (uV*sec)	% Area
1	Peak1	bb	4.772	0.449	534	0.025
2	Peak2	bb	5.924	0.557	638	0.030
3	Peak3	bv	9.448	0.888	401	0.019
4	Peak4	vb	9.528	0.896	579	0.028
5	Urea IDN 6893	BB	10.636	1.000	2100455	99.875
6	Peak6	bb	13.942	1.311	470	0.022

Crystallization of 4-chloro-7-methoxyquinoline-6-carboxamide (*2*)

2
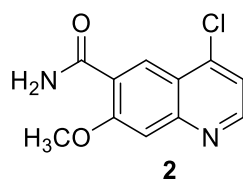
4-Chloro-7-methoxyquinoline-6-carboxamide **2** (99.1 g, 418.7 mmol) was suspended in 595 mL (6V) of *N*-methyl-2-pyrrolidinone (NMP) and the mixture was heated at 100°C to have complete dissolution. After cooling to room temperature in about 2 h, the suspension was stirred at room temperature for 20 h and at 0°c for 1h. The suspension was filtered, the solid washed with cold (0÷5°C) NMP (50 mL, 0.5V) and triturated three times with water (300 mL, 3V). Finally, the filtrate was dried at 60°C in vacuum for 24h affording a light brown solid (83.2 g, y = 83%). Typical HPLC purity (λ = 250 nm) is 99.61%.
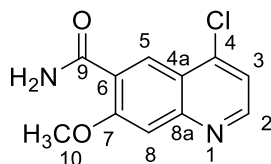
^{1}H-NMR (DMSO-d$_6$) δ (ppm): 4.03 (3H, s, OMe), 7.57 (1H, s, H$_8$), 7.63 (1H, d, *J* = 6.6 Hz, H$_3$), 7.79 (1H, brs, NH), 7.88 (1H, brs, NH), 8.48(1H,s, H$_5$), 8.8 (1H, d, *J* = 6.6 Hz, H$_2$).
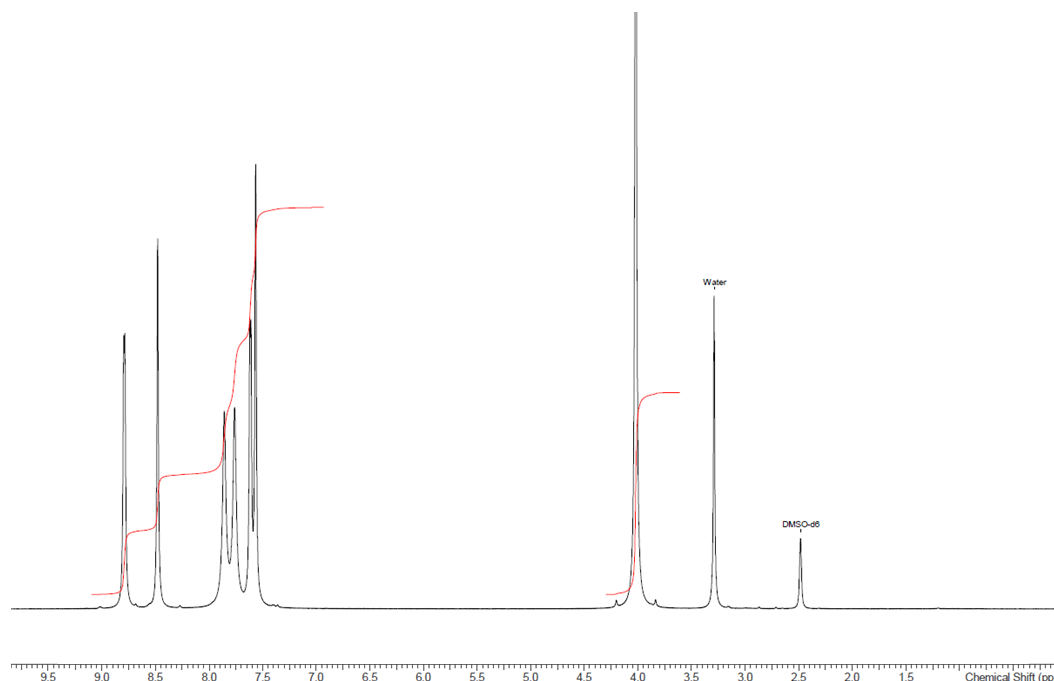
^{13}C-NMR (DMSO-d$_6$) δ (ppm): 56.8 (C$_{10}$), 109.0 (C$_8$), 120.2 (C$_6$), 120.5 (C$_3$), 126.7 (C$_5$), 127.7 (C$_{4a}$), 142.1 (C$_{8a}$), 151.2 (C$_4$), 152.6 (C$_2$), 158.6 (C$_7$), 166.1 (C$_9$).
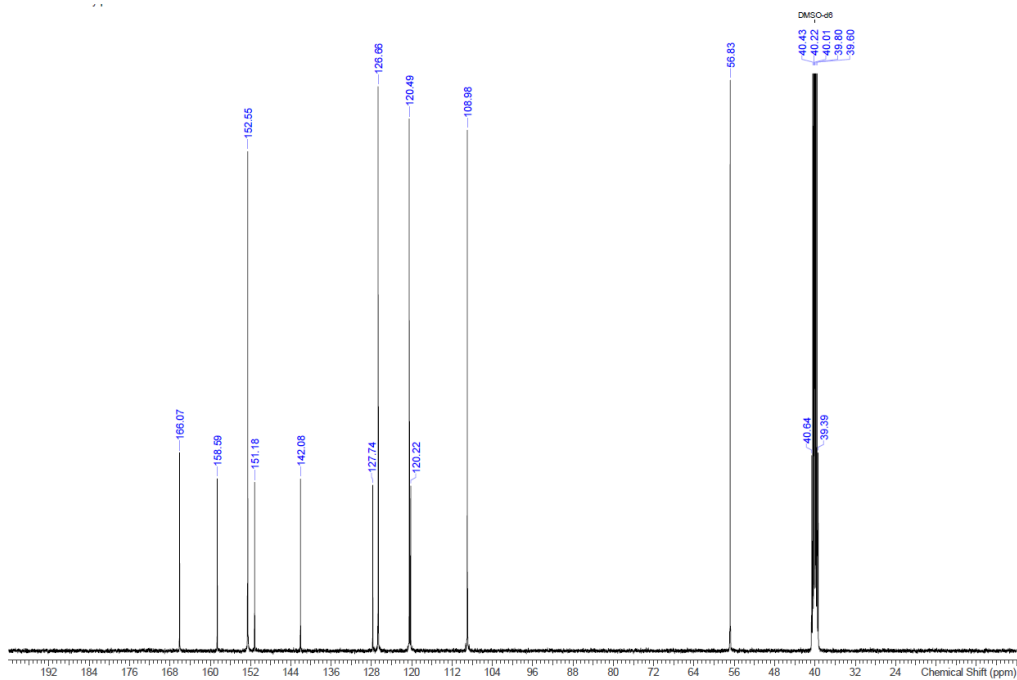
ESI-MS: m/z = 237.0 [M+H]$^+$

HPLC chromatogram (λ = 250 nm): Typical HPLC purity is 99.61%.
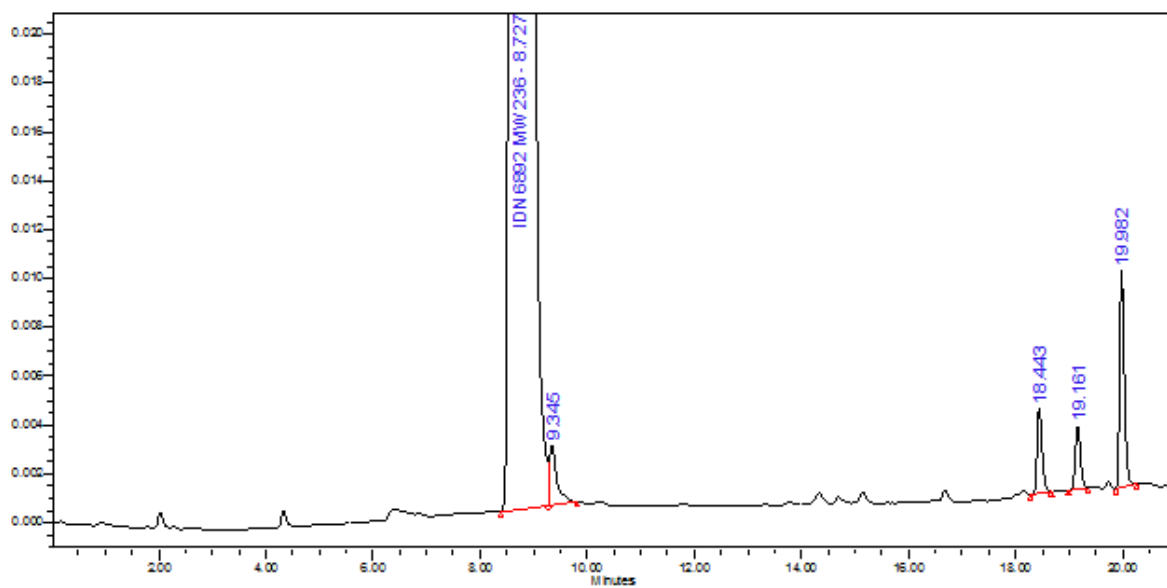
	Name	Int Type	RT (min.)	Area (uV*sec)	% Area
1	IDN 6892 MW 236	bv	8.727	29882513	99.605
2		vb	9.345	24638	0.082
3		bb	18.443	22547	0.075
4		bb	19.161	17433	0.058
5		bb	19.982	53965	0.180

Syntheisis of Lenvatinib free base (1)
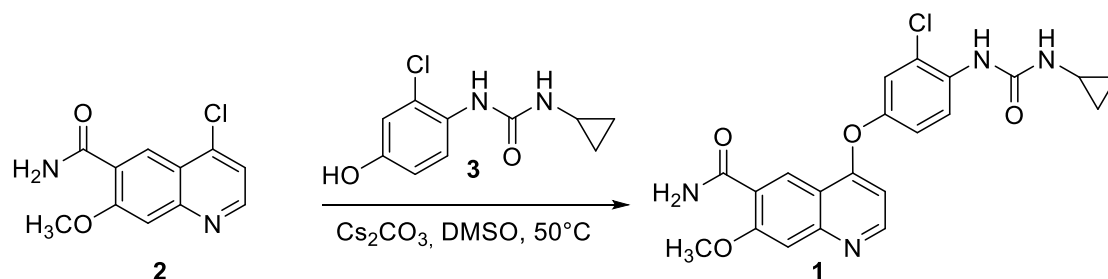
Recrystallized 4-chloro-7-methoxyquinoline-6-carboxamide **2** (32.0 g, 135.2 mmol, 1 eq.) and 1-(2-chloro-4-hydroxyphenyl)-3-cyclopropylurea **3** (61.30 g, 270.4 mmol, 2 eq.) were suspended under nitrogen in DMSO (192 mL, 6V). Cesium carbonate (88.11 g, 270.4 mmol, 2 eq) was added and the mixture was stirred at 50°C under a nitrogen atmosphere for 24 h. The reaction was monitored by HPLC (sample preparation: 50 μL of reaction mixture were diluted in 1 mL of 1 M solution of HCl in 1:1 CH_3CN : H_2O and 50 μL of this solution further diluted in 1 mL H_2O with 0.3%TFA). Typically after 24 h, A% of 4-chloro-7-methoxyquinoline-6-carboxamide **2** vs A% of Lenvatinib **1** satisfies the limit of NMT 3%. The reaction mixture was cooled to room temperature and the precipitated inorganic salts were filtered off. The filtrate was diluted with DMSO (192 mL, 6V) and water (192 mL, 6V) added drop-wise in 40 min. to precipitate the product. The obtained suspension was stirred at room temperature for 1h. The precipitated crystals were filtered, washed with 1:1 DMSO : H_2O (64 mL, 2V) and triturated three times with water (130 mL, 4V). The solid was dried at 60°C under vacuum for 24 h to afford a light brown/grey solid (53.25 g, 124.7 mmol, y = 92%).

Crystallization of Lenvatinib

Crude Lenvatinib (53.25 g, 124.7 mmol) was dissolved in DMSO (450 mL, 5V) at 70°C. The solution was cooled to room temperature and dichloromethane (1350 mL, 15V) was added in 15 min. The mixture was stirred at room temperature for 16 h and at 0÷5°C for 1 h. The suspension was filtered, the solid washed with 1:3 DMSO : DCM (106 mL, 2V) and triturated with pure DCM (210 mL, 4 V) three times. The solid was dried at 60°C under vacuum for 20 h to yield Lenvatinib fee base **(1)** (38.87 g, 91.03 mmol, crystallization yield = 73 %, total yield = 67%).
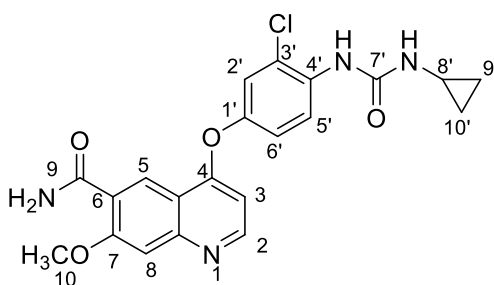
<u>1H-NMR</u> (DMSO-d$_6$) δ (ppm): 0.43 (2H, m, H$_{9'A}$,H$_{10'A}$), 0.65 (2H, m, H$_{9'B}$,H$_{10'B}$), 2.40 (3H, s, CH$_3$SO$_3$H), 2.57 (1H, m, H$_{8'}$), 4.08 (3H, s, OMe), 6.46 (1H, d, J = 6.6 Hz, H$_3$), 7.18 (1H, brs, NH$_{8'}$), 7.25(1H, dd, J = 2.8 Hz, J = 9.1 Hz, H$_{6'}$), 7.62 (1H,d, J = 2.8 Hz, H$_{2'}$), 7.70(1H, s, H$_8$), 7.87 (1H, s, NH$_9$), 7.94 (1H, s, NH$_9$), 8.06 (1h, s, NH$_{4'}$), 8.34 (1H, d, J = 9.1 Hz, H$_{5'}$), 8.71 (1H, s, H$_5$), 8.63 (1H, d, J = 6.6 Hz, H$_2$)

<u>HPLC chromatogram (λ = 309 nm) of crude Lenvatinib:</u> Typical HPLC purity is 97.34 %.
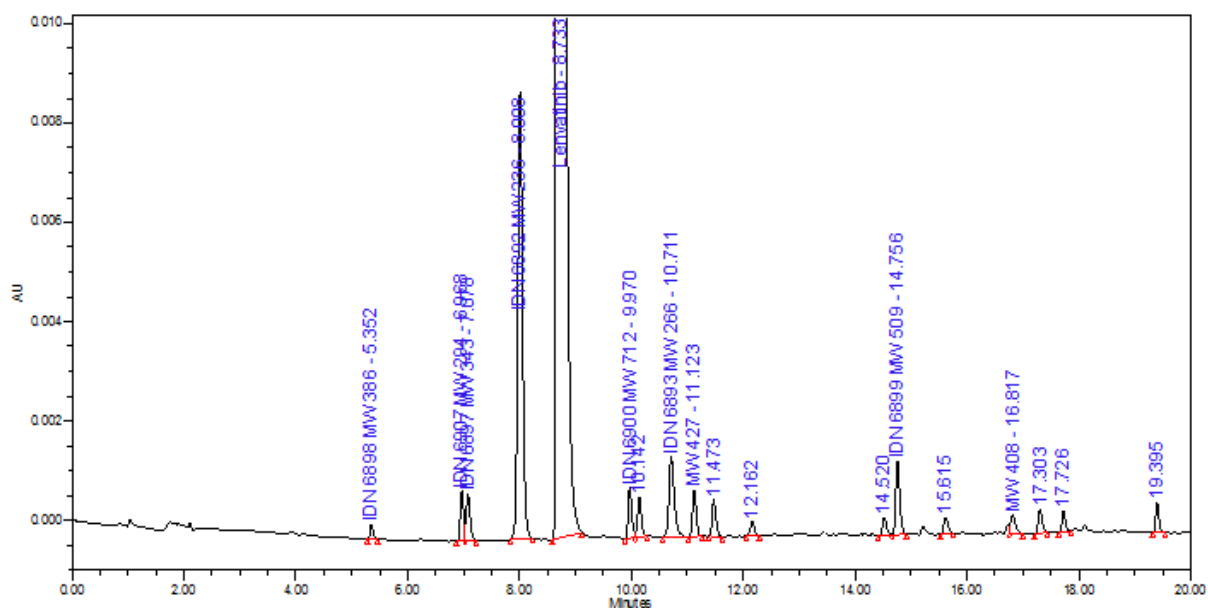
	Name	Int Type	RT (min.)	RRT	Area (uV*sec)	% Area
1	IDN 6898 MW 386	Bb	5.352	0.613	1137	0.027
2	IDN 6907 MW 294	bV	6.968	0.798	4126	0.099
3	IDN 6897 MW 343	VB	7.078	0.810	4308	0.104
4	IDN 6892 MW 236	BB	8.008	0.917	52474	1.263
5	Lenvatinib	BB	8.733	1.000	4042991	97.339
6	IDN 6900 MW 712	BV	9.970	1.142	4686	0.113
7		VB	10.142		3740	0.090
8	IDN 6893 MW 266	BB	10.711	1.226	11189	0.269
9	MW 427	BB	11.123	1.274	4348	0.105
10		BB	11.473		3534	0.085
11		BB	12.162		1350	0.033
12		BV	14.520		1888	0.045
13	IDN 6899 MW 509	VB	14.756	1.690	7272	0.175
14		BB	15.615		1844	0.044
15	MW 408	vb	16.817	1.926	2277	0.055
16		BB	17.303		2331	0.056
17		BB	17.726		1723	0.041

<u>HPLC chromatogram (λ = 309 nm) of recrystallized Lenvatinib</u>: Typical HPLC purity is 99.60%.
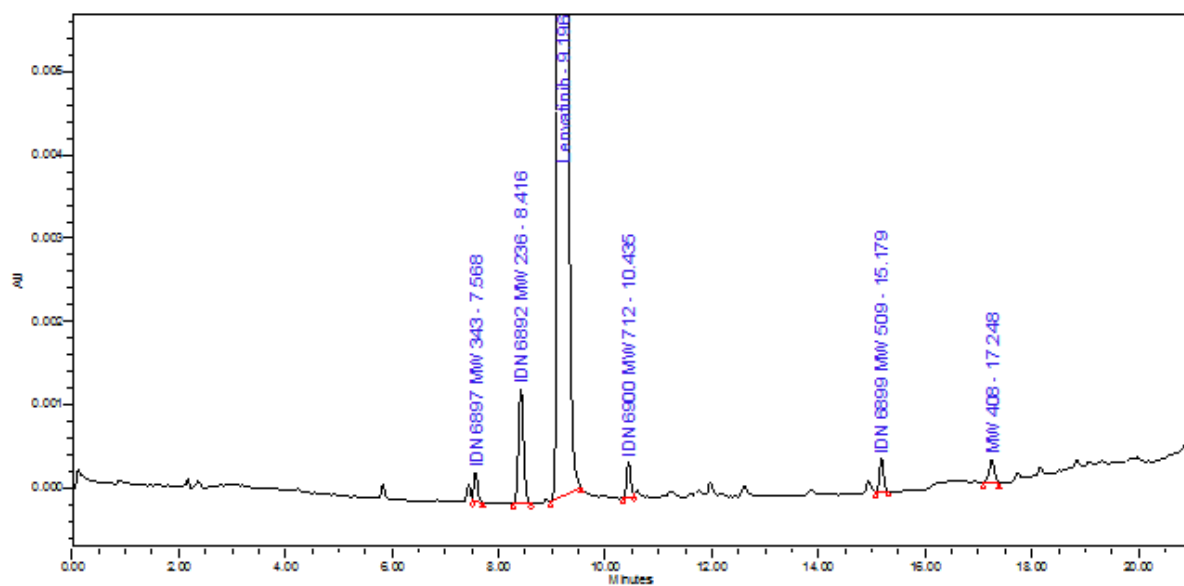
	Name	Int Type	RT (min.)	RRT	Area (uV*sec)	% Area
1	IDN 6897 MW 343	vB	7.568	0.823	1751	0.041
2	IDN 6892 MW 236	BB	8.416	0.915	9002	0.211
3	Lenvatinib	BB	9.196	1.000	4258630	99.599
4	IDN 6900 MW 712	Bv	10.435	1.135	2242	0.052
5	IDN 6899 MW 509	BB	15.179	1.651	2139	0.050
6	MW 408	bB	17.248	1.876	1990	0.047

Mesylate salt of Lenvatinib – ACA-1-HT-dry (**1.Ms**)
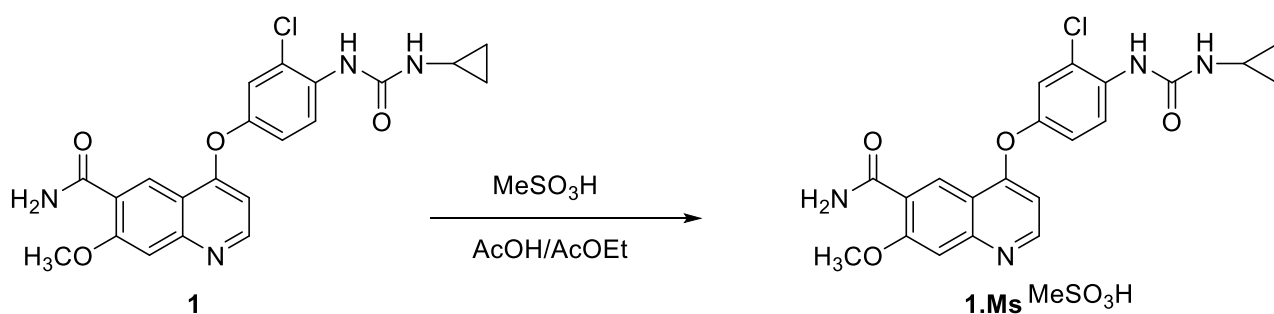
Half amount of Lenvatinib free base **1** (19.0 g, 44.5 mmol, 0.5 eq.) was suspended in acetic acid (103 mL, 2.7 V) and half amount of methanesulfonic acid (2.9 mL, 4.3 g, 44.5 mmol, 0.5 eq.) diluted in acetic acid (6 mL, 0.15V) was added. The mixture was stirred at room temperature for 10' until it was homogeneous. The second half of Lenvatinib free base **1** (19.0 g, 44.5 mmol, 0.5 eq.) and methanesulfonic acid (2.9 mL, 4.28 g, 44.5 mmol, 0.5 eq.) diluted in acetic acid (6 mL, 0.15V) were added. The mixture was stirred at 60°C for 30 min. to reach complete dissolution. The solution was passed through Whatman 0.2 μm filter and heated again at 60°C. Ethyl acetate (38 mL, 1 V) was added drop-wise and the temperature was decreased to 40°C in 1 h. The mixture was stirred at 40°C for 16 h (at this temperature a white solid precipitates) and at room temperature for 1 h. Ethyl acetate (190 mL, 5 V) was added and the suspension was filtered. The solid was washed with 1:2 Acetic Acid : EtOAc (76 mL, 2V). The wet solid (ACA-1 form ,114 g) was dried at 70°C in vacuum for 120 h to afford the title compound **1.Ms** (34.8 g, 66.7 mmol, y = 75%). Typical HPLC purity is 99.72 %.
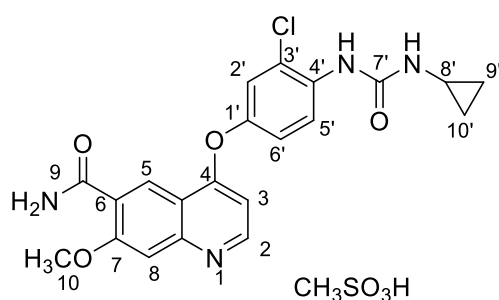
<u>^{1}H-NMR</u> (DMSO-d_6) δ (ppm): 0.43 (2H, m, $H_{9'A}$,$H_{10'A}$), 0.65 (2H, m, $H_{9'B}$,$H_{10'B}$), 2.40 (3H, s, CH_3SO_3H), 2.57 (1H, m, $H_{8'}$), 4.08 (3H, s, OMe), 6.96 (1H, d, J = 6.6 Hz, H_3), 7.25 (1H, brs, $NH_{8'}$), 7.35 (1H, dd, J = 2.8 Hz, J = 9.1 Hz, $H_{6'}$), 7.62 (1H,d, J = 2.8 Hz, $H_{2'}$), 7.70(1H, s, H_8), 7.87 (1H, s, NH_9), 7.94 (1H, s, NH_9), 8.06 (1h, s, $NH_{4'}$), 8.34 (1H, d, J = 9.1 Hz, $H_{5'}$), 8.71 (1H, s, H_5), 8.97 (1H, d, J = 6.6 Hz, H_2).

Spectrum N.12493
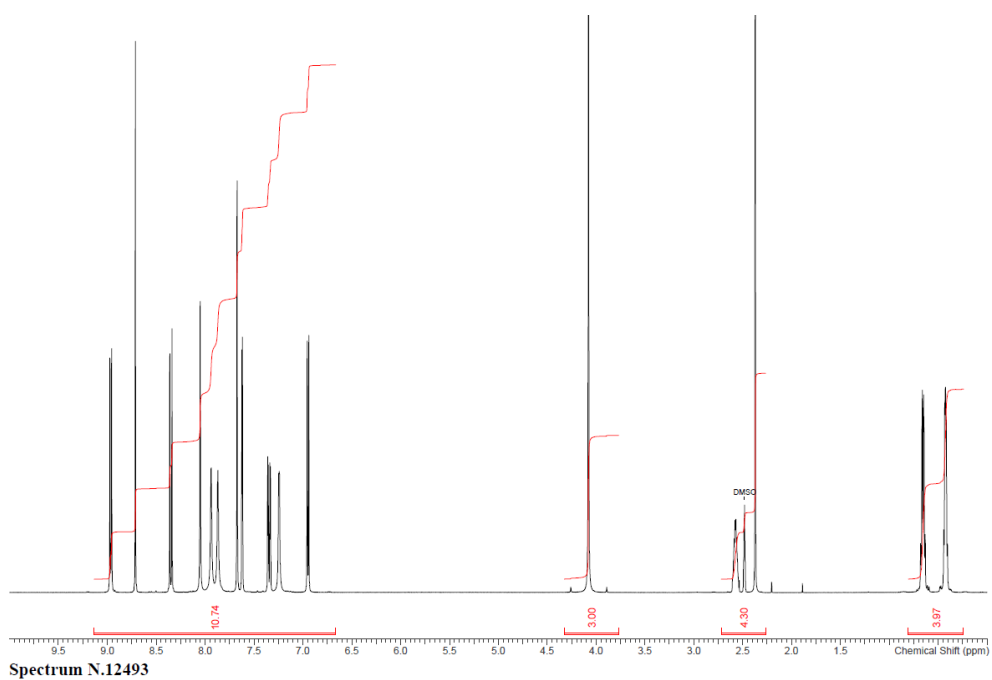
^{13}C-NMR (DMSO-d$_6$) δ (ppm): 6.7 (C$_{9'}$, C$_{10'}$), 22.8 (C$_{8'}$), 40.3 (CH$_3$SO$_3$H), 57.4 (C$_{10}$), 101.2 (C$_8$), 104.1 (C$_3$), 114.5 (C$_{4a}$), 120.8 (C$_{6'}$), 122.2 (C$_{5'}$), 122.4 (C$_{2'}$), 122.6 (C$_{3'}$), 125.7 (C$_5$), 128.7 (C$_6$), 136.3 (C$_{4'}$), 143.1 (C$_{8a}$), 146.5 (C$_{1'}$), 148.5 (C$_2$), 155.9 (C$_{7'}$), 161.6 (C$_7$), 165.3 (C$_9$), 167.8 (C$_4$).

Spectrum N.12493
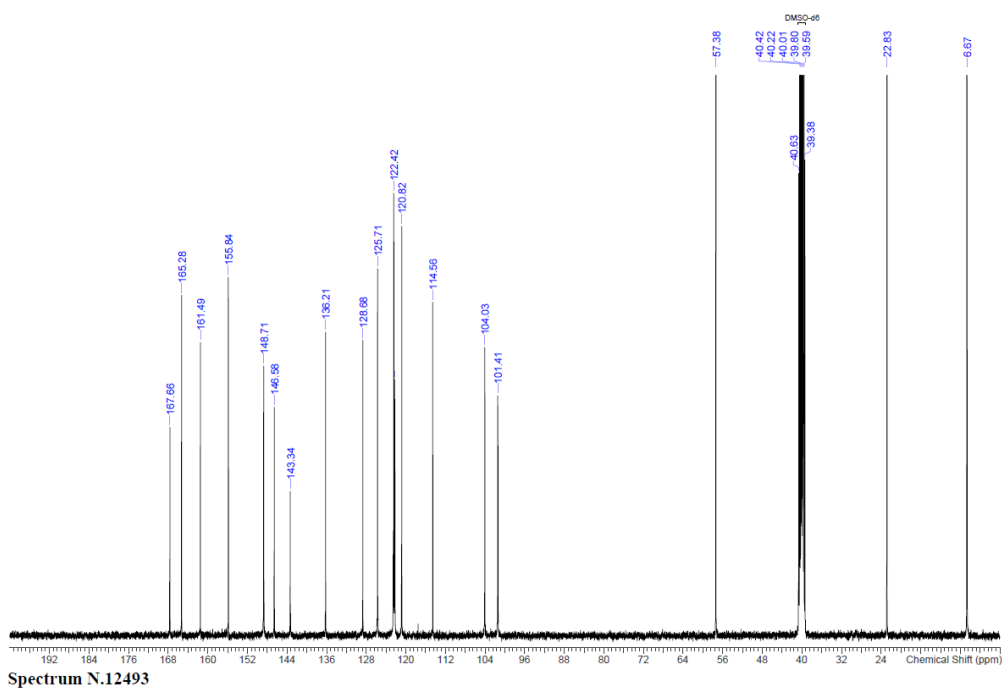
ESI-MS: m/z = 427.4 [M+H]$^+$; 852.9 [2M+H]$^+$.

HPLC chromatogram (λ = 309 nm): Typical HPLC purity is 99.72 %.
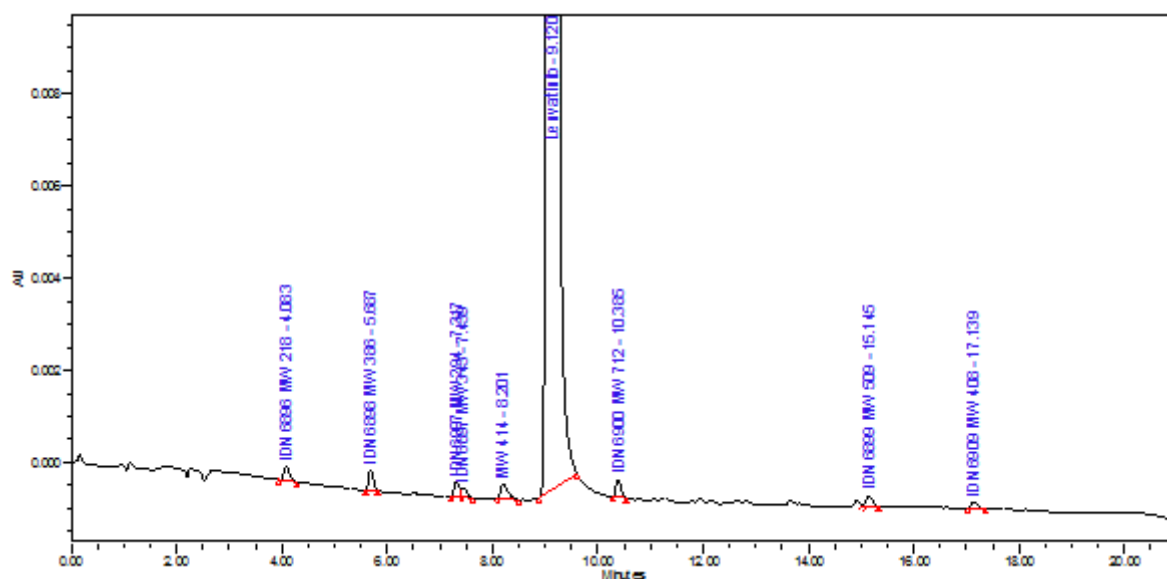
	Name	Int Type	RT (min.)	RRT	Area (uV*sec)	% Area
1	IDN 6896 MW218	bb	4.083	0.448	2247	0.036
2	IDN 6898 MW386	BB	5.687	0.624	2864	0.046
3	IDN 6907 MW294	bv	7.317	0.802	2424	0.039
4	IDN 6897 MW343	vb	7.439	0.816	1431	0.023
5	MW414	VB	8.201	0.899	3201	0.051
6	Lenvatinib	BB	9.120	1.000	6210906	99.715
7	IDN 6900 MW712	Bb	10.385	1.139	2531	0.041
8	IDN 6899 MW509	vb	15.145	1.661	1839	0.030
9	IDN 6909 MW408	bb	17.139	1.879	1217	0.020

Residual solvents

- Acetic acid = 398 ppm

- EtOAc < LOD

- DCM < LOD

- DMSO < LOD

Water content (KF) = 0.56%.

Solid state characterization was reported in experimental section of Chapter 2.

Synthesis of 4-Hydroxy-7-methoxyquinoline-6-carboxamide (*10*)
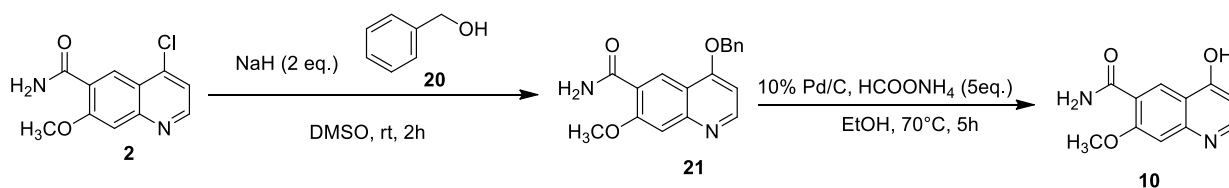
After dissolving benzylic alcohol **20** (216 mg, 2 mmol, 2 eq) in dry DMSO (6 mL, 28V), NaH (96 mg, 4 mmol, 4 eq.) was gradually added at room temperature and the mixture was stirred for 30 minutes. 4-chloro-7-methoxy-6-quinolinecarboxamide **2** (236 mg, 1 mmol, 1 eq.) was added, and the resulting mixture was stirred at room temperature for additionally 2 h. The reaction solution was distributed between ethyl acetate and water, and the organic layer was dried over Na_2SO_4. After filtration, the solvent was removed under vacuum and the solid was washed with diethyl ether to afford intermediate **21** (191 mg, 0.62 mmol, y =62%).

10% Pd/C (60mg, 0.06 mmol, 0.1 eq.) was added to a stirred solution of 4-benziloxy-7-methoxyquinoline-6-carboxamide **21** (191 mg, 0.62 mmol, 1 eq.) and sublimed ammonium formate (195 mg, 3.1 mmol, 5 eq) in ethanol (15 mL, 78 V). The mixture was stirred at 70°C for 5 h. Then, upon cooling to room temperature, the catalyst was filtered off and the solvent was removed under vacuum. The crude was purified by flash chromatography column (8:2 EtOAc : methanol) to yield product **10** (97 mg, 0.45 mmol, y = 75 %, overall yield = 47%).
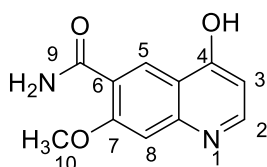
^{1}H-NMR (DMSO-d_6) δ (ppm): 3.93 (3H, s, OCH_3), 5.95 (1H, d, J = 7.6 Hz, H_3), 7.02 (1H, s, H_8), 7.50 (1H, brs, NH), 7.61 (1H, brs, NH), 7.81 (1H, d, J = 6.6 Hz, H_2), 8.51 (1H, s, H_5).
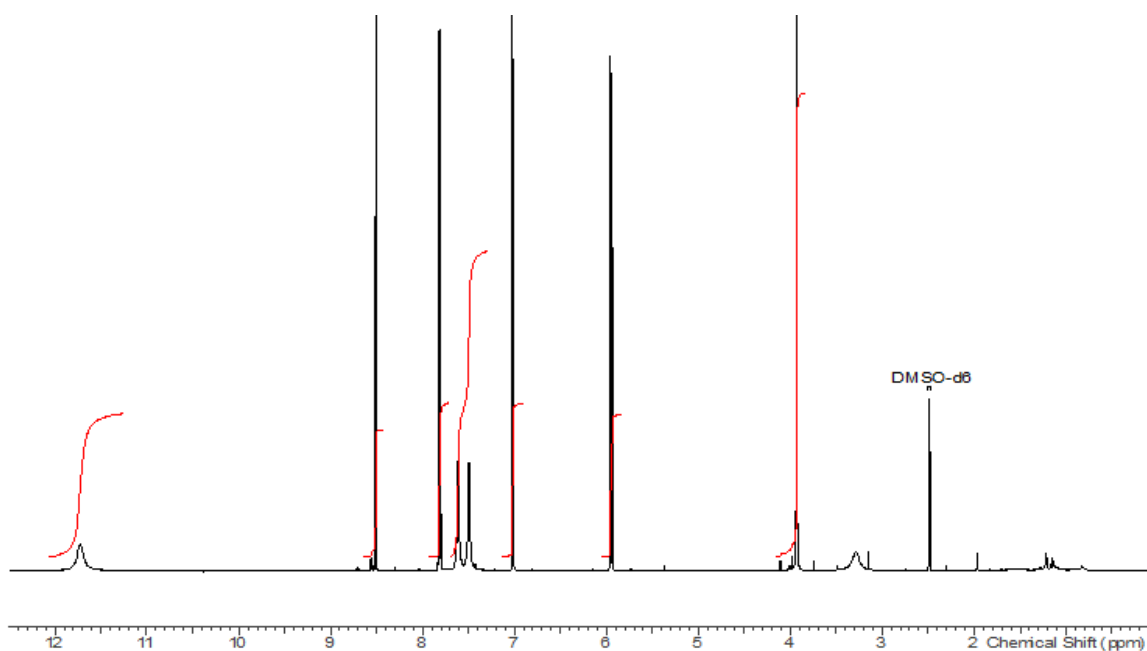
^{13}C-NMR (DMSO-d$_6$) δ (ppm): 56.5 (C$_{10}$), 99.2 (C$_8$), 109.6 (C$_3$), 119.8 (C$_6$), 120.5 (C$_{4a}$), 129.7 (C$_5$), 140.0 (C$_2$), 143.6 (C$_{8a}$), 159.8 (C$_7$), 166.1 (C$_4$), 177.0 (C$_9$).
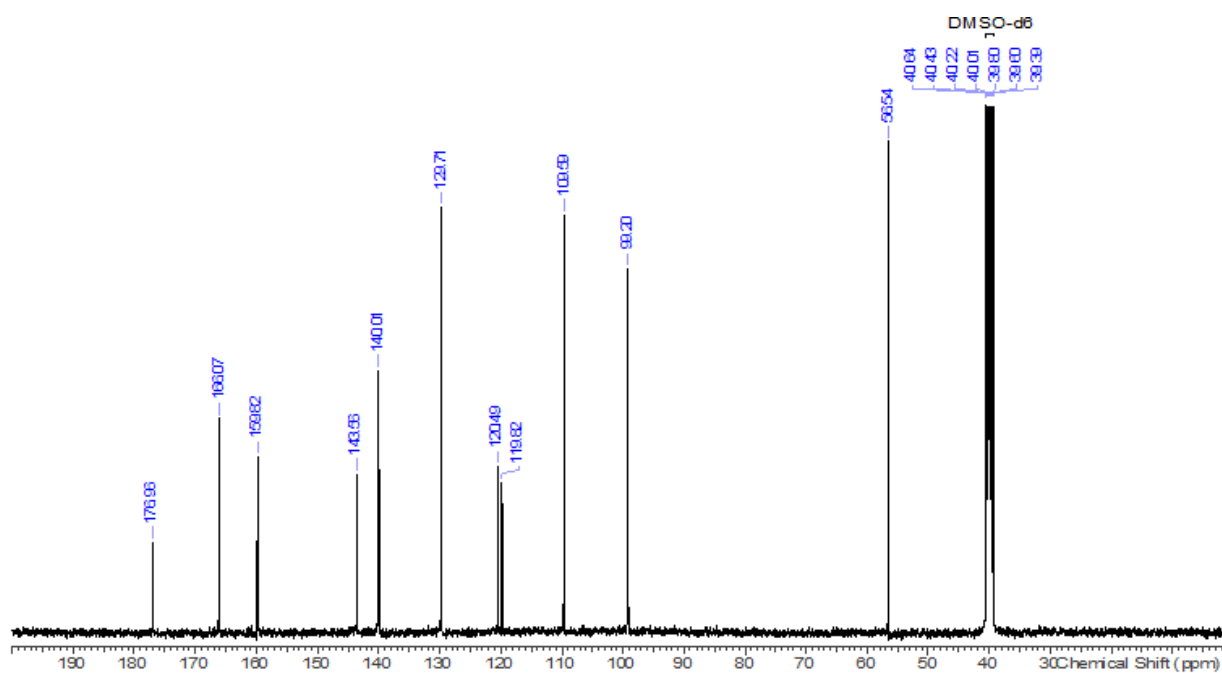
ESI-MS: m/z = 219.1 [M+H]$^+$

FTIR ν (cm^{-1}): 3437 stretching N-H; 3291-3060 stretching C-H (aromatic); 2952-2824 Stretching C-H (aliphatic); 1686 stretching C=O (amide I band); 1635-1560 stretching C=C (aromatic); 1412-1350

bending O-H; 1216-1200 Stretching C-O (aryl alcohol); 1052 Stretching C-O-C (symmetric, aryl alkyl ether).
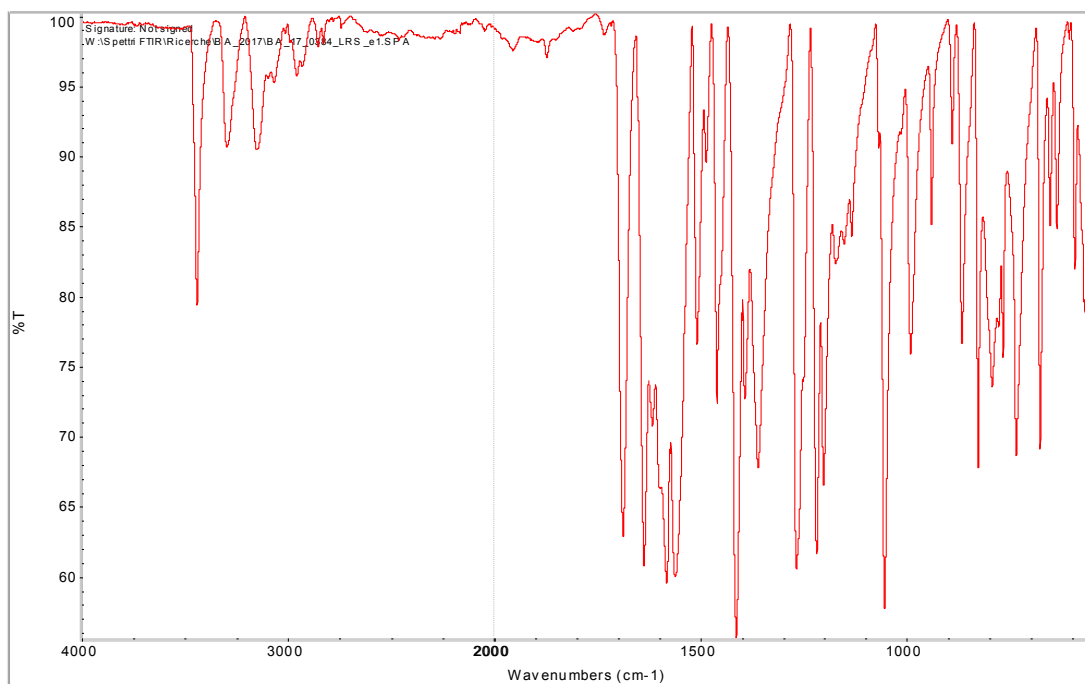
HPLC purity is 99.82%.
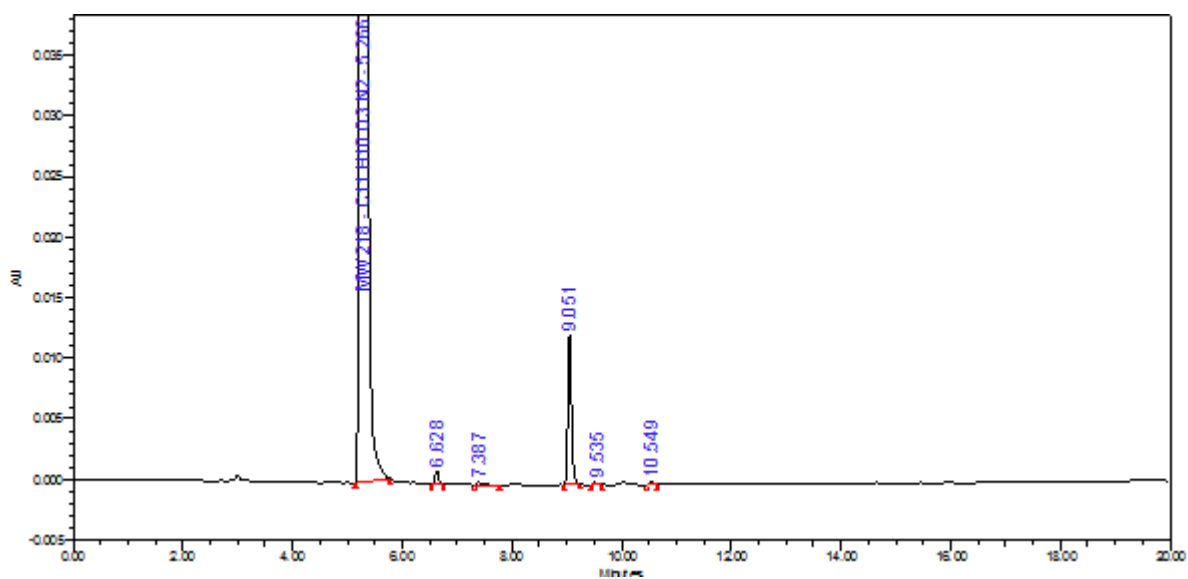
	Name	Int Type	RT (min.)	Area (uV*sec)	% Area
1	MW 218 - C11 H10 O3 N2	BB	5.266	5495779	98.811
2		BB	6.628	4917	0.088
3		BB	7.387	1123	0.020
4		BB	9.051	57249	1.029
5		BB	9.535	1475	0.027
6		BB	10.549	1369	0.025

Synthesis of 7-methoxy-4-phenoxyquinoline-6-carboxamide (*12*)
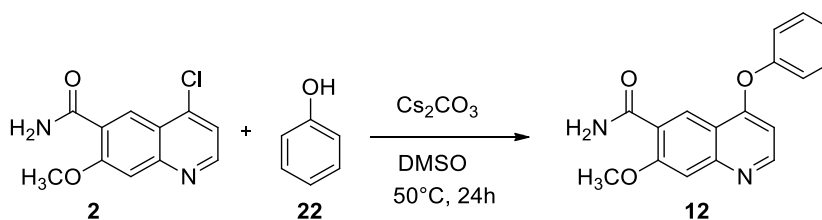
4-Chloro-7-methoxy-6-quinolinecarboxamide **2** (2.00 g, 8.45 mmol, 1 eq.) and phenol **1** (1.59 g, 14.90 mmol, 2 eq.) were suspended in DMSO (12 mL, 6V) under nitrogen flow. Cesium carbonate (5.50 g, 14.90 mmol, 2 eq.) was added and the mixture was stirred at 50°C for 24 h in nitrogen atmosphere. Upon cooling to room temperature, water (12 mL, 6V) was added and the resulting suspension filtered. The solid was washed with 1:1 DMSO : water (4 mL), triturated three times with water (8mL, 4V) and dried in vacuum at 60°C for 20 h to yield product **12** (2.38 g, 8.10 mmol, y = 95%).
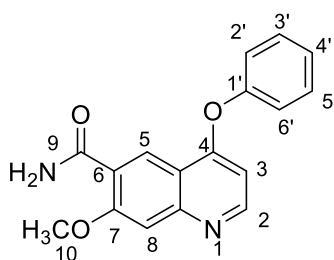
^{1}H-NMR (DMSO-d$_6$) δ (ppm): 4.02 (3H, s, OCH$_3$), 6.46 (1H, d, *J* = 5.9 Hz, H$_3$), 7.29 (2H, d, *J* = 7.7 Hz, H$_{2'}$, H$_{6'}$), 7.34 (1H, t, *J* = 7.4 Hz, H$_{4'}$), 7.51 (1H, s, H$_8$), 7.53 (2H, d, *J* = 7.6 Hz, H$_{3'}$, H$_{5'}$), 7.71 (1H, brs, NH), 7.84 (1H, brs, NH), 8.65 (1H, d, *J* = 5.3 Hz, H$_2$), 8.51 (1H, s, H$_5$).
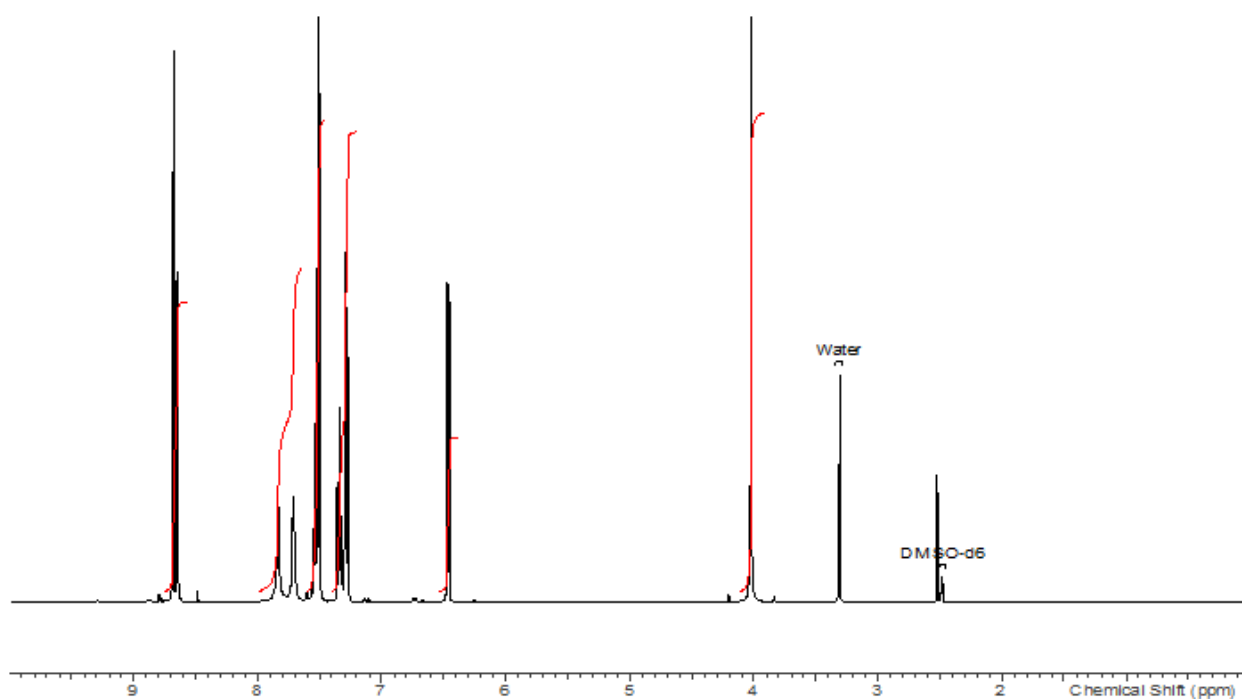
¹³C-NMR (DMSO-d₆) δ (ppm): 56.6 (C$_{10}$), 103.7 (C$_3$), 108.4 (C$_8$), 115.1 (C$_{4a}$), 121.3 (C$_{6'}$, C$_{2'}$), 125.2 (C$_5$), 125.5 (C$_6$), 126.3 (C$_{4'}$), 131.0 (C$_{3'}$, C$_{5'}$), 152.2 (C$_{8a}$), 153.7 (C$_2$), 154.2 (C$_{1'}$), 158.5 (C$_7$), 161.9 (C$_4$), 166.2 (C$_9$).
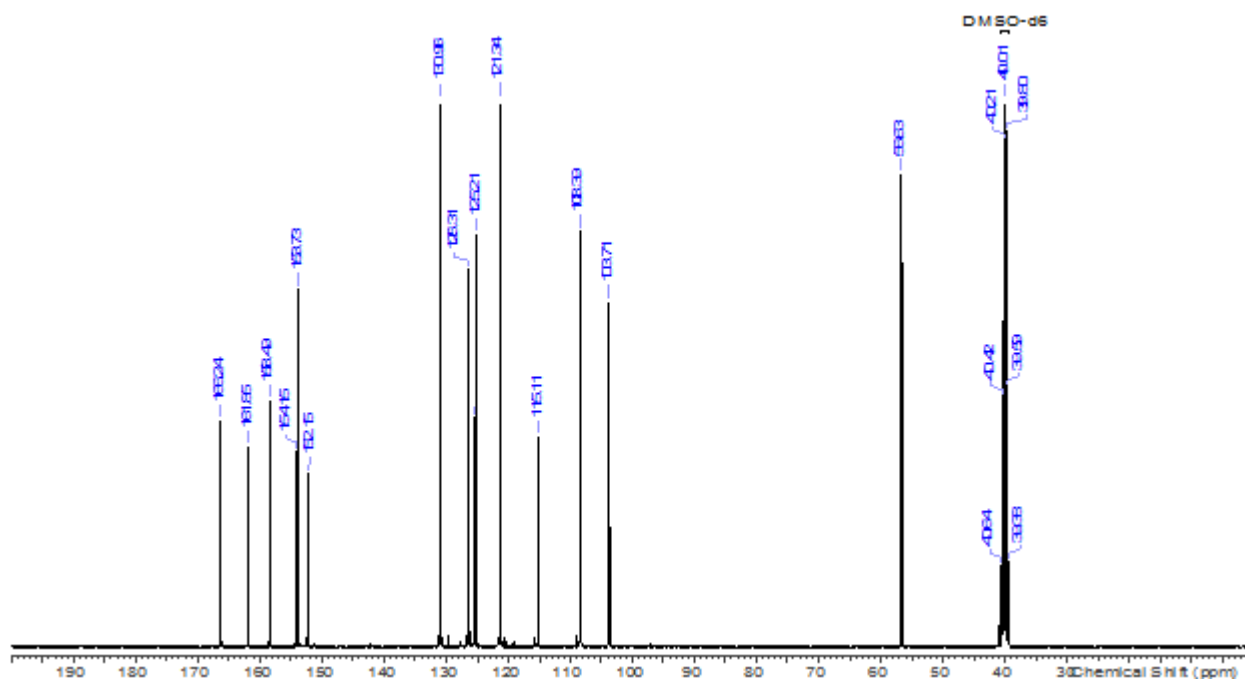
ESI-MS: m/z = 295.2 [M+H]$^+$

<u>FTIR</u> v (cm^{-1}): 3460 stretching N-H; 3200-3000 stretching C-H (aromatic); 2950-2800 Stretching C-H (aliphatic); 1670 stretching C=O (amide I band); 1615-1562 stretching C=C (aromatic); 1230-1204 Stretching C-O (aryl ether); 1057 Stretching C-O-C (symmetric, aryl alkyl ether).
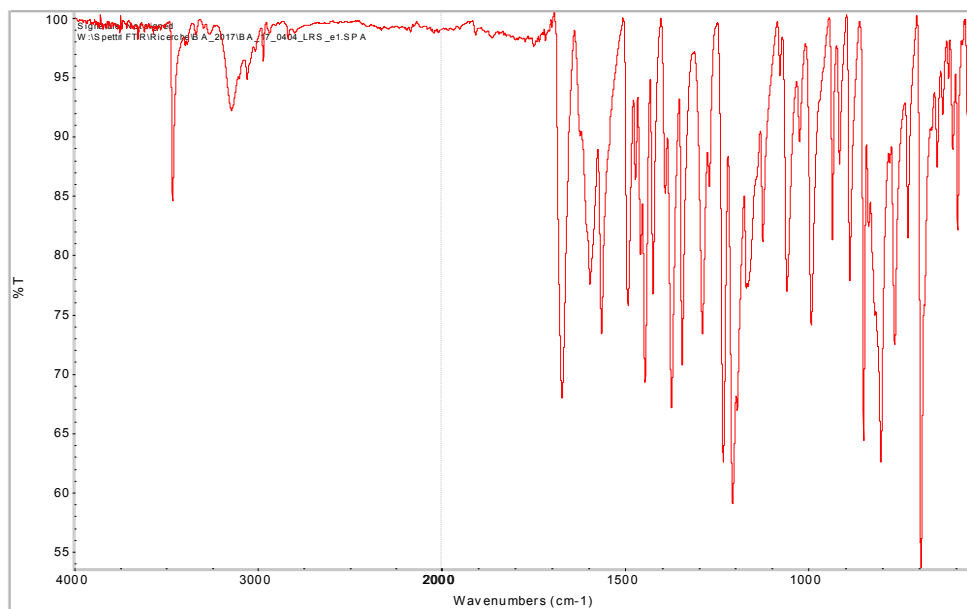
<u>HPLC</u> purity is 99.02%.
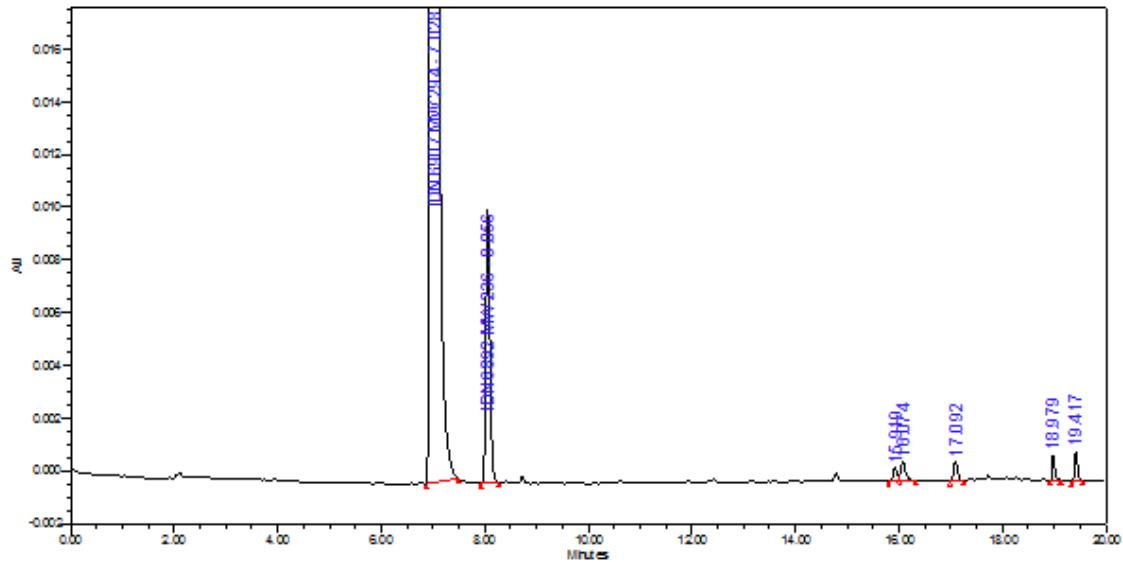
	Name	Int Type	RT (min.)	Area (uV*sec)	% Area
1	IDN 6907 MW 294	BB	7.028	7721562	99.017
2	IDN 6892 MW 236	BB	8.056	57255	0.734
3		bv	15.919	2768	0.035
4		vb	16.074	4739	0.061
5		bb	17.092	3966	0.051
6		bb	18.979	3480	0.045
7		bb	19.417	4478	0.057

Synthesis of 4-(4-amino-3-chlorophenoxy)-7-methoxyquinoline-6-carboxamide (13)
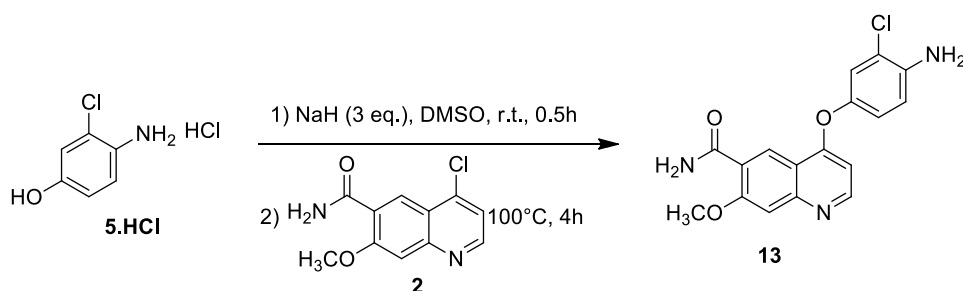
After dissolving 4-amino-3-chlorophenol hydrochloride salt **1.HCl** (360 mg, 2 mmol, 2 eq) in dry DMSO (6 mL, 17V), NaH (144 mg, 6 mmol, 6 eq.) was gradually added at room temperature and the mixture was stirred for 30 minutes. 4-chloro-7-methoxy-6-quinolinecarboxamide **2** (236 mg, 1 mmol, 1 eq.) was added and the resulting mixture was stirred at 100°C for 4 h. Upon cooling to room temperature, the reaction solution was distributed between ethyl acetate and water. The organic layer was washed with water, saturated brine and dried over Na_2SO_4. The solid was filtered off and the solvent removed under vacuum. The crude product was purified by flash chromatography (95:5 EtOAc. MeOH) to afford the desired product **13** (189 mg, 0.55 mmol, y = 55%).
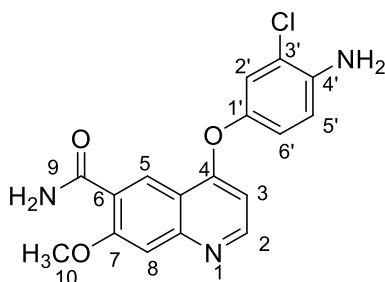
$\underline{^1\text{H-NMR}}$ (DMSO-d_6) δ (ppm): 4.01 (3H, s, OCH₃), 5.41 (2H, s, NH₂-4') 6.44 (1H, d, J = 5.3 Hz, H₃), 6.90 (1H, d, J = 8.7 Hz, H₅'), 7.00 (1H, dd, J = 8.7 Hz, J = 2.6 Hz, H₆'), 7.22 (1H, d, J = 2.6 Hz, H₂'), 7.48 (1H, s, H₈), 7.67 (1H, brs, NH-9), 7.82 (1H, brs, NH-9), 8.62 (1H, d, J = 5.3 Hz, H₂), 8.66 (1H, s, H₅).
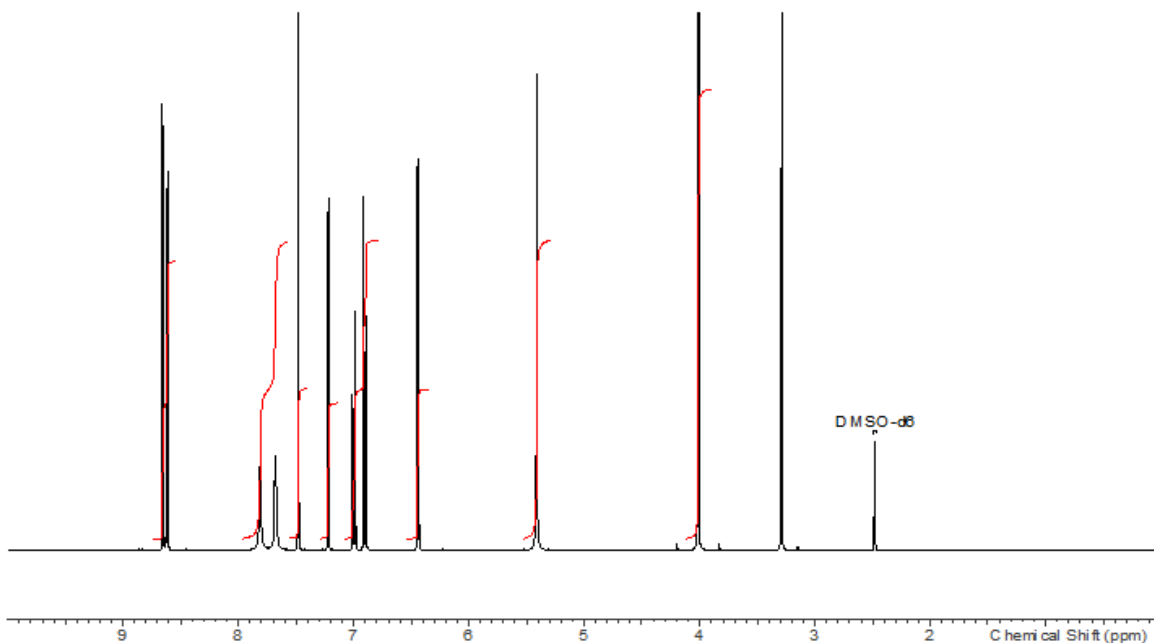
¹³C-NMR (DMSO-d$_6$) δ (ppm): 56.6 (C$_{10}$), 102.9 (C$_3$), 108.3 (C$_8$), 114.8 (C$_{4a}$), 116.5 (C$_{5'}$), 117.6 (C$_{3'}$), 121.4 (C$_{6'}$), 122.3 (C$_{2'}$), 125.3 (C$_5$, C$_6$), 143.5 (C$_{4'}$, C$_{1'}$), 152.0 (C$_{8a}$), 153.8 (C$_2$), 158.4 (C$_7$), 162.7 (C$_4$), 166.2 (C$_9$).
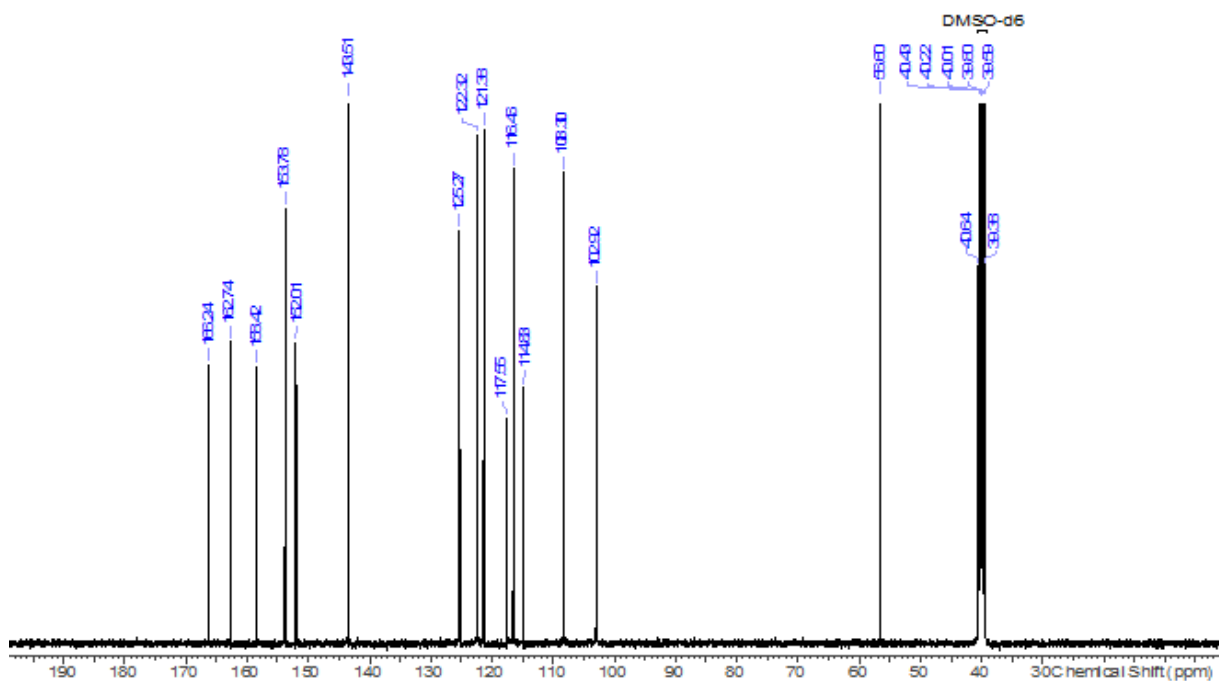
ESI-MS: m/z = 344.1 [M+H]$^+$; 366.1 [M+Na]$^+$; 709.1 [2M+H]$^+$

<u>FTIR</u> v (cm^{-1}): 3497 stretching N-H; 3378 stretching N-H; 3212-3013 stretching C-H (aromatic); 1651 stretching C=O (amide I band); 1621-1564 stretching C=C (aromatic); 1275-1050 Stretching C-O-C (ethers).
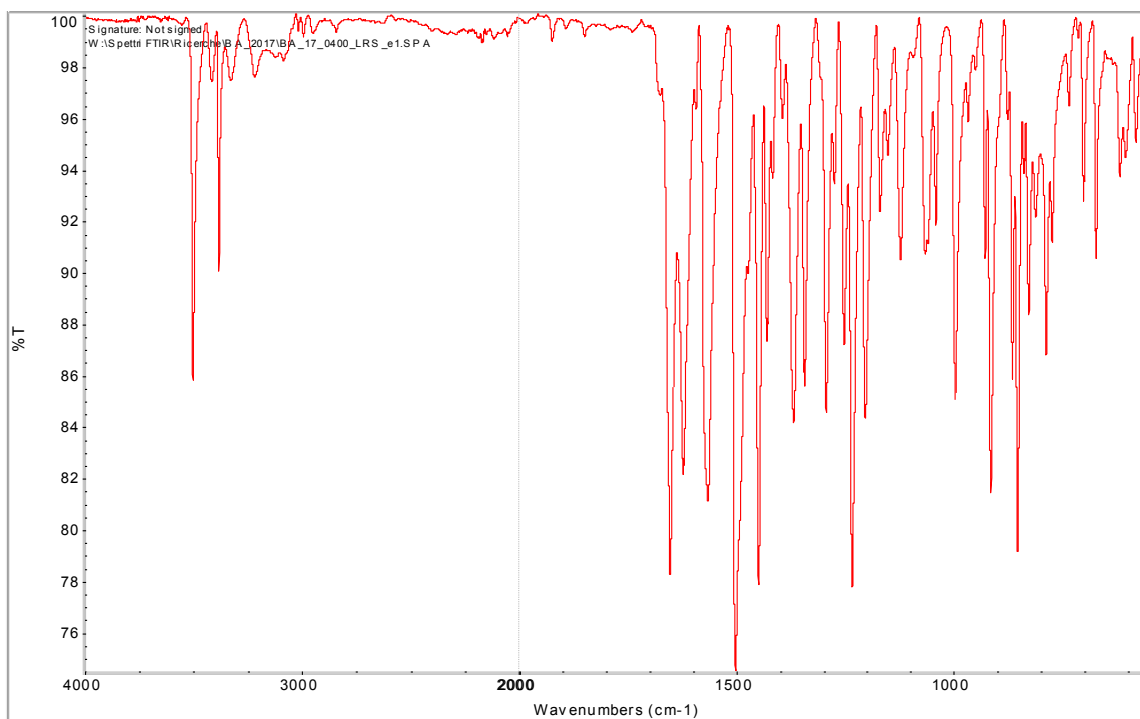
<u>HPLC</u> purity is 99.96%.
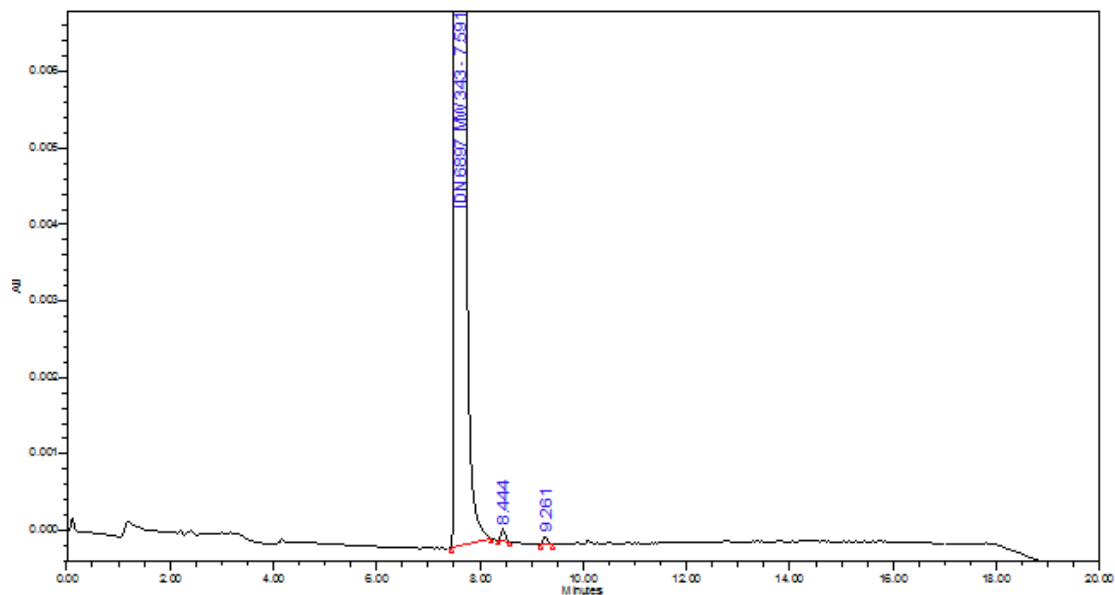
	Name	Int Type	RT (min.)	Area (uV*sec)	% Area
1	IDN 6897 MW 343	Bb	7.591	3327648	99.955
2		bb	8.444	913	0.027
3		bb	9.261	571	0.017

Synthesis of 4-(3-Chloro-4-ureidophenoxy)-7-methoxyquinoline-6-carboxamide (11)
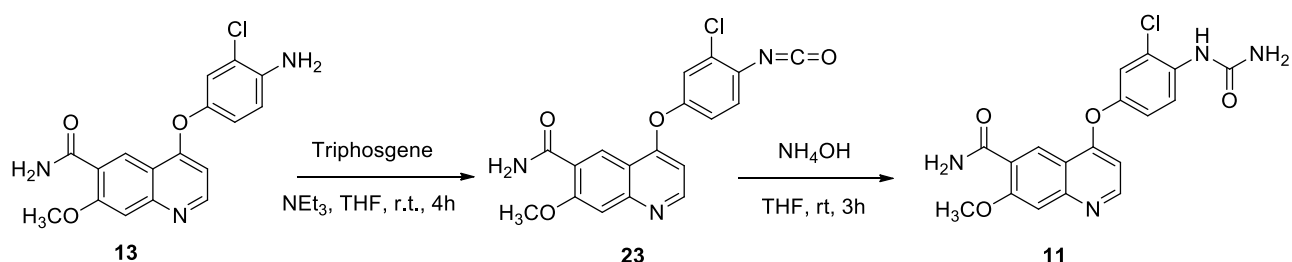
A solution of triphosgene (45 mg, 0.15 mmol, 1 eq.), in dry THF (150 μL), was prepared inside a dried round bottom flask equipped with nitrogen flow. Then, a solution of **13** (50 mg, 0.15mmol, 1 eq.) in dry THF (1.2mL) was dropped inside the flask. The reaction mixture was cooled down to 0°C and triethylamine (44 μL, 32 mg. 0.315 mmol, 2.1eq) was added dropwise. After starting the addition, a white gas was released. The reaction was stirred for 4 h at room temperature (monitoring by FTIR). The volatiles were evaporated to yield isocyanate **23**, which was used without further purification.

Isocyanate **23** was dissolved in THF (1mL) and 30% ammonium hydroxide solution (35 mg, 0.3 mmol, 2 eq.) was added dropwise. The reaction was stirred for 3 h at room temperature until FTIR showed complete consumption of isocyanate (2265 cm-1 band disappeared). The volatiles were removed at reduced pressure and the crude was purified by column chromatography (95:5 EtOAc : MeOH, Rf = 0.25) to yield compound **11** (19 mg, 0.05 mmol, y = 32%).
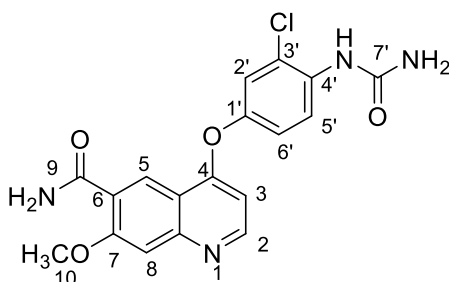
^{1}H-NMR (DMSO-d_6) δ (ppm): 4.02 (3H, s, OCH$_3$), 6.44 (2H, brs, NH$_2$-7') 6.53 (1H, d, J = 5.3 Hz, H$_3$), 7.21 (1H, dd, J = 9.1 Hz, J = 2.8 Hz, H$_{6'}$), 7.46 (1H, d, J = 2.8 Hz, H$_{2'}$), 7.51 (1H, s, H$_8$), 7.69 (1H, brs, NH-9), 7.82 (1H, brs, NH-9), 8.20 (1H, brs, NH-4'), 8.24 (1H, d, J = 9.1 Hz, H$_{5'}$) 8.65 (1H, s, H$_5$), 8.66 (1H, d, J = 4.9 Hz, H$_2$).
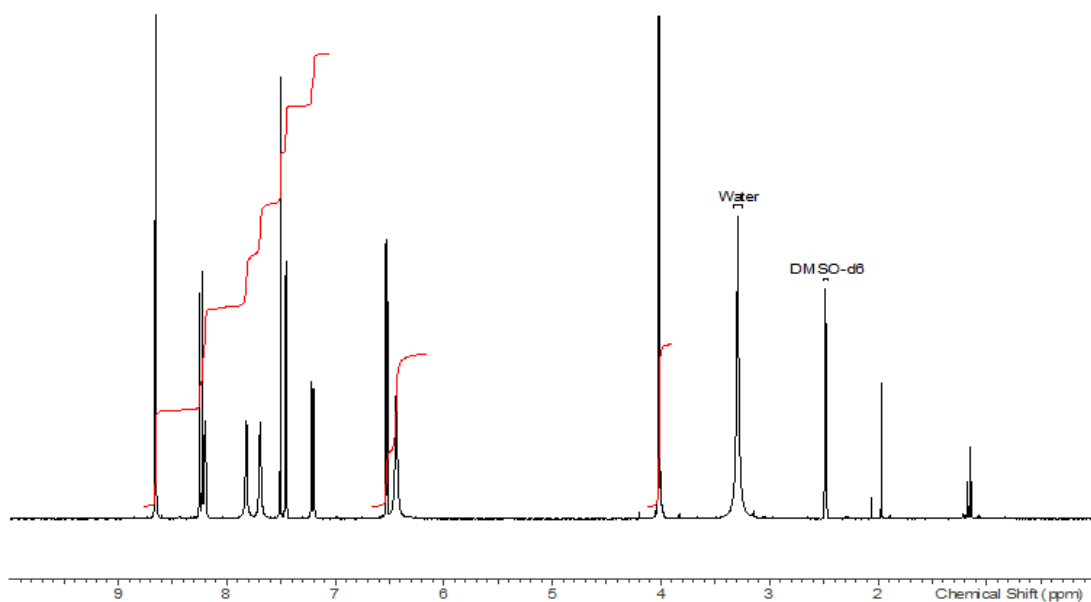
^{13}C-NMR (DMSO-d$_6$) δ (ppm): 56.7 (C$_{10}$), 103.5 (C$_3$), 108.2 (C$_8$), 114.9 (C$_{4a}$), 120.7 (C$_{6'}$), 122.3 (C$_{2'}$), 122.7 (C$_{3'}$), 122.8 (C$_{5'}$), 125.2 (C$_5$), 125.6 (C$_6$), 135.3 (C$_{4'}$), 148.0 (C$_{1'}$), 151.9 (C$_{8a}$), 153.7 (C$_2$), 156.1 (C$_{7'}$), 158.6 (C$_7$), 162.1 (C$_4$), 166.2 (C$_9$).
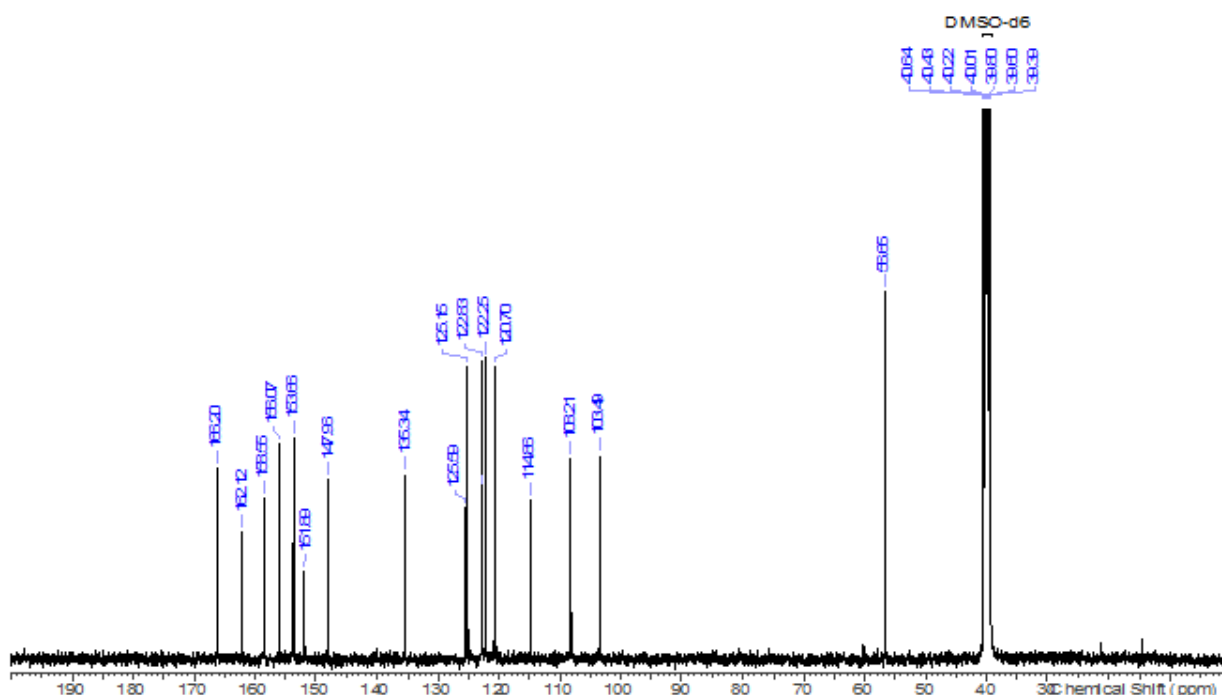
ESI-MS: m/z = 387.1 [M+H]$^+$; 409.1 [M+Na]$^+$; 425.0 [M+K]$^+$; 795.1 [2M+H]$^+$

FTIR ν (cm^{-1}): 3441 stretching N-H (amide); 3265-3184 stretching N-H (urea); 1667 stretching C=O (amide I band + urea); 1582-1522 stretching C=C (aromatic); 1275-1050 Stretching C-O-C (ethers).
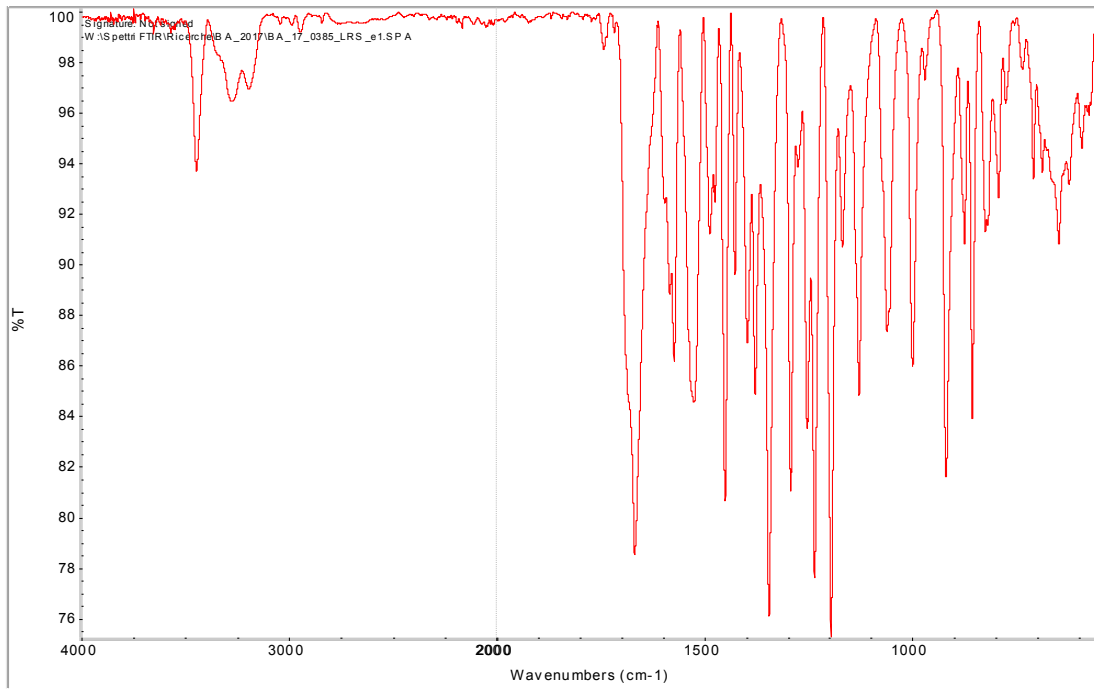
HPLC purity is 97.35%.
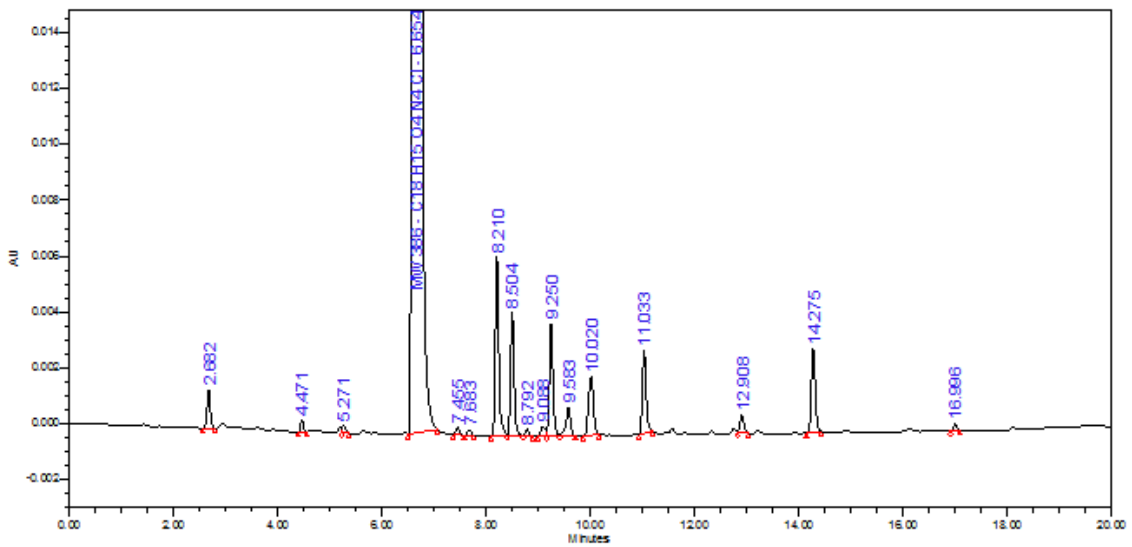
	Name	Int Type	RT (min.)	Area (uV*sec)	% Area
1		bb	2.682	6145	0.117
2		BB	4.471	1585	0.030
3		VB	5.271	1178	0.022
4	MW 386 - C18 H15 O4 N4 Cl	BB	6.654	5123050	97.355
5		BB	7.455	1284	0.024
6		BB	7.683	985	0.019
7		BV	8.210	29830	0.567
8		VV	8.504	20899	0.397
9		VB	8.792	1042	0.020

	Name	Int Type	RT (min.)	Area (uV*sec)	% Area
10		BV	9.088	2110	0.040
11		VV	9.250	19321	0.367
12		VB	9.583	6009	0.114
13		BB	10.020	13949	0.265
14		BB	11.033	14677	0.279
15		VB	12.908	3449	0.066
16		BB	14.275	15743	0.299
17		BB	16.996	970	0.018

Synthesis of 4,4'-(((carbonylbis(azanediyl))bis(3-chloro-4,1-phenylene))bis(oxy)) bis-(7-methoxyquinoline-6-carboxamide) (14)
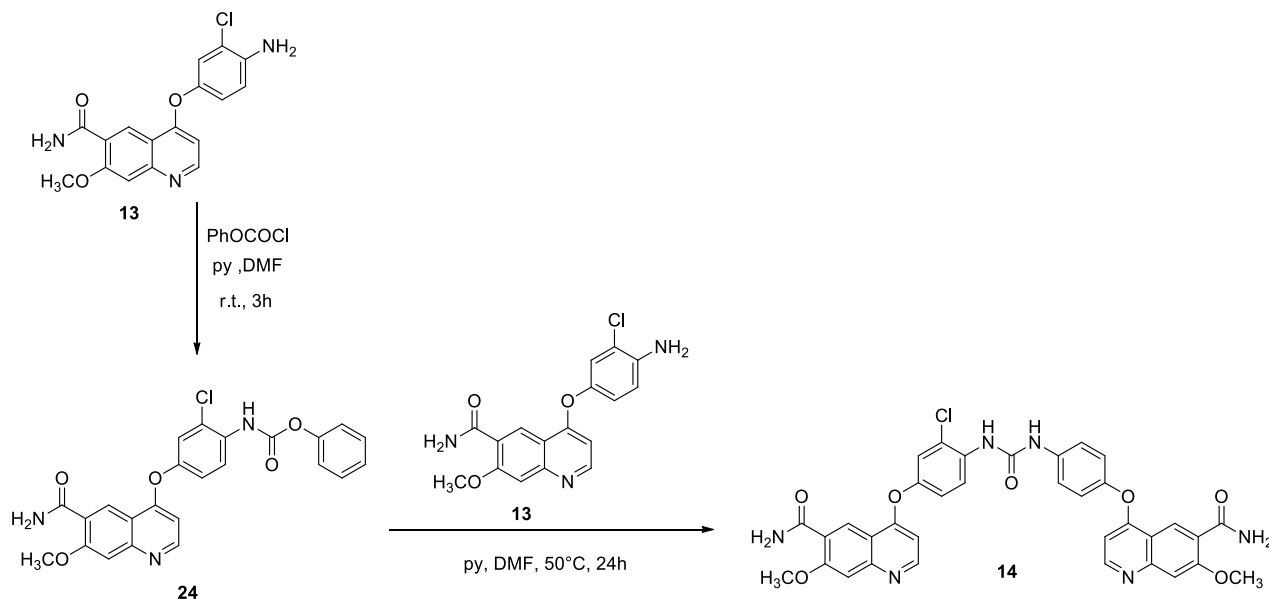
Comppound **13** (100 mg, 0.29 mmol, 1 eq.) was charged into a 25mL flask, previously dried. Then, DMF (1mL, 10V) was added under nitrogen and the solution cooled down to 0°C. Pyridine (70 μL, 69 mg, 0.87 mmol, 3eq) and phenyl chloroformate (65.6 μL, 52 mg, 0.32 mmol, 1.1 eq.) were added. The reaction mixture was stirred at room temperature for 3 h (TLC: 100% ethyl acetate R_f (product **24)** = 0.24). The reaction mixture was diluted with water the aqueous phase was extracted 3 times with ethyl acetate. The combined organic layers were washed with water and brine, dried over sodium sulfate, filtered and the volatiles evaporated under vacuum to yield compound **24** as pale orange solid (114 mg, 0.25 mmol, y =85%).

Compound **24** (50 mg, 0.11mmol, 1 eq.) was added into a dried round bottom flask equipped with nitrogen inlet and dissolved in dry DMF (1mL, 20V). Amine **13** (38 mg, 10.11 mmol, 1 eq.) was added and then pyridine (17.7 μL, 17.4 mg, 0.22mmol, 2 eq.) was added drop-wise. The reaction mixture was stirred at 50°C for 24 h until TLC (9:1 EtOAc: MeOH) showed the complete consumption of phenyl carbamate **24** and the presence of product **14** (R_f = 0.1). The volatiles were removed at reduced pressure and the crude was purified by column chromatography (300 mL of 100%EtOAc + 700 mL of 9:1 EtOAc : MeOH) to yield product **14** (27 mg, 0.04 mmol, y = 35%, overall yield = 30%).
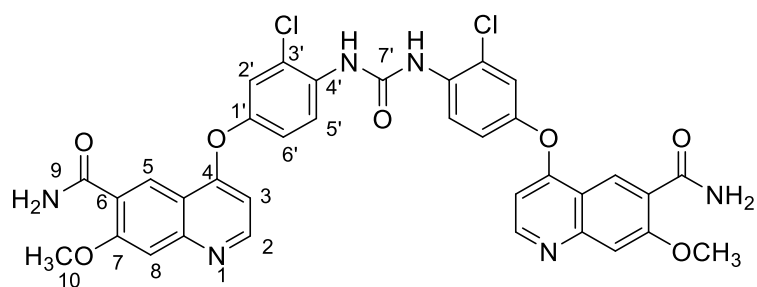
¹H-NMR (DMSO-d₆) δ (ppm): 4.02 (6H, s, 2xOCH₃), 6.58 (2H, d, *J* = 5.1 Hz, 2xH₃), 7.29 (2H, dd, *J* = 9.0 Hz, *J* = 2.2 Hz, 2xH₆′), 7.51 (2H, s, 2xH₈), 7.55 (2H, d, *J* = 2.1 Hz, 2xH₂′), 7.70 (2H, brs, 2xNH-9), 7.83 (2H, brs, 2xNH-9), 8.19 (2H, d, *J* = 8.9 Hz, 2xH₅′) 8.66 (2H, s, 2xH₅), 8.68 (2H, d, *J* = 5.5 Hz, 2xH₂), 9.22 (2H, brs, 2xNH-7′).
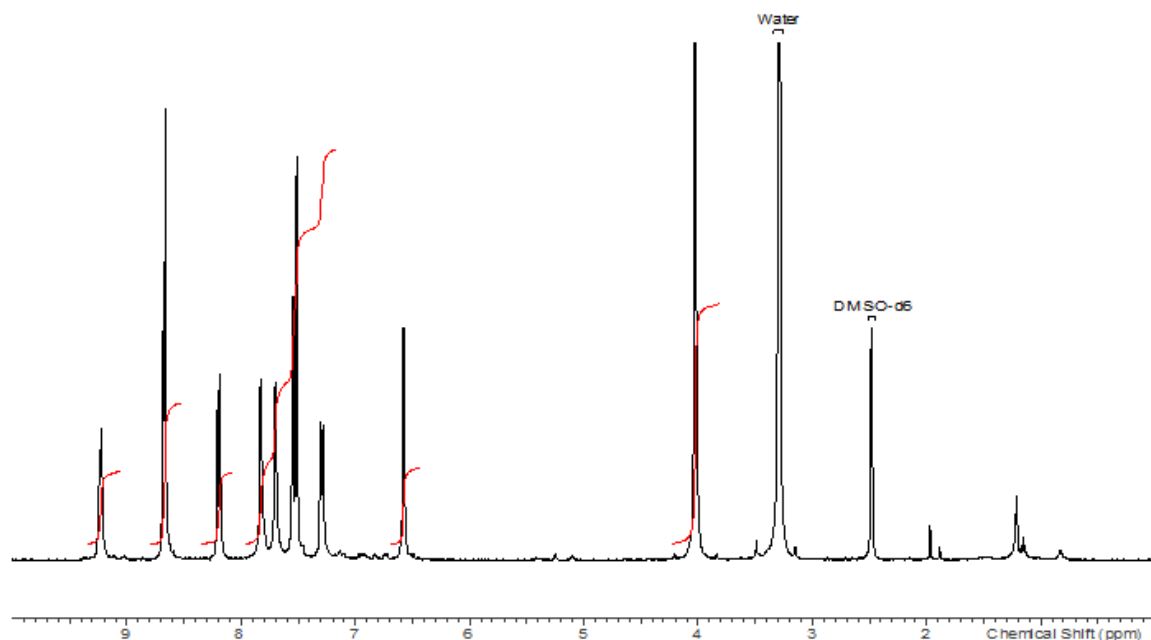
¹³C-NMR (DMSO-d₆) δ (ppm): 56.7 (2xC₁₀), 103.8 (2xC₃), 108.4 (2xC₈), 114.9 (2xC₄ₐ), 120.8 (2xC₆′), 122.5 (2xC₂′), 124.3 (2xC₅′), 124.4 (2xC₃′), 125.2 (2xC₅), 125.6 (2xC₆), 134.2 (2xC₄′), 149.2 (2xC₁′), 152.1 (2xC₈ₐ), 152.9 (C₇′), 153.8 (2xC₂), 158.5 (2xC₇), 161.8 (2xC₄), 166.2 (2xC₉).
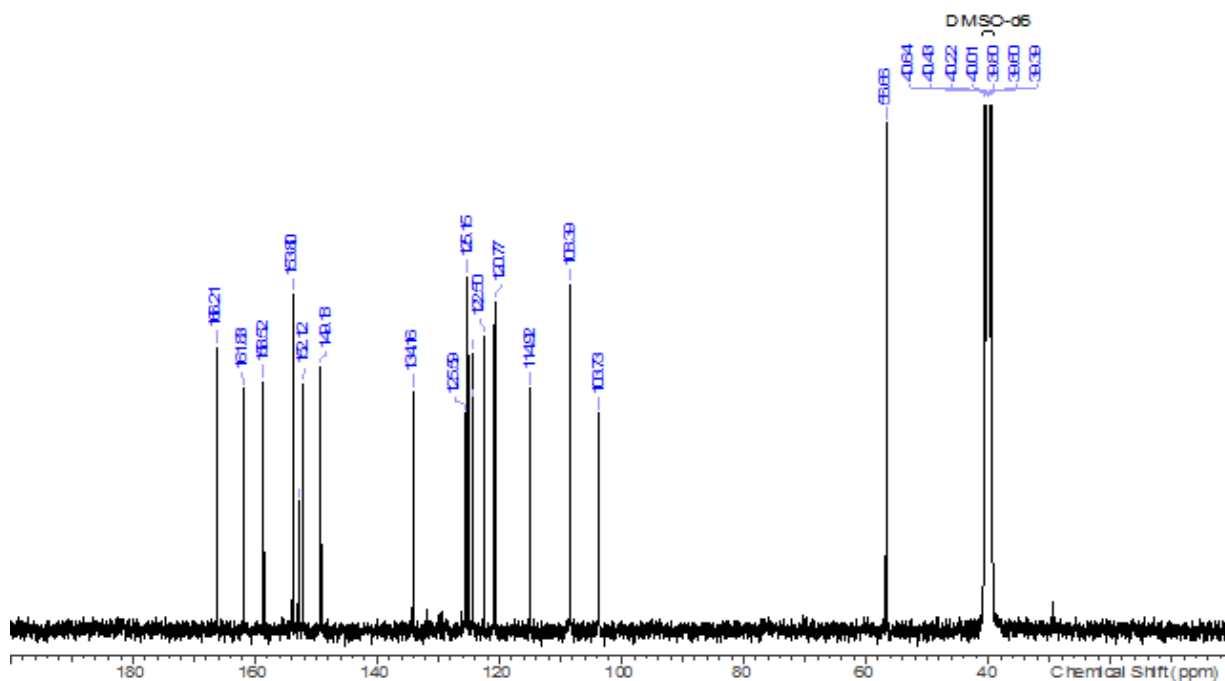
ESI-MS: m/z = 357.1 [M+2H]$^{2+}$; 713.1 [M+H]$^{+}$; 735.1 [M+Na]$^{+}$; 751.1 [M+K]$^{+}$

FTIR v (cm^{-1}): 3455 stretching N-H (amide); 3301-3157 stretching N-H (urea); 1679 stretching C=O (amide I band + urea); 1647-1532 stretching C=C (aromatic); 1231 stretching C-O (aryl alcohol); 1064 Stretching Ar-Cl
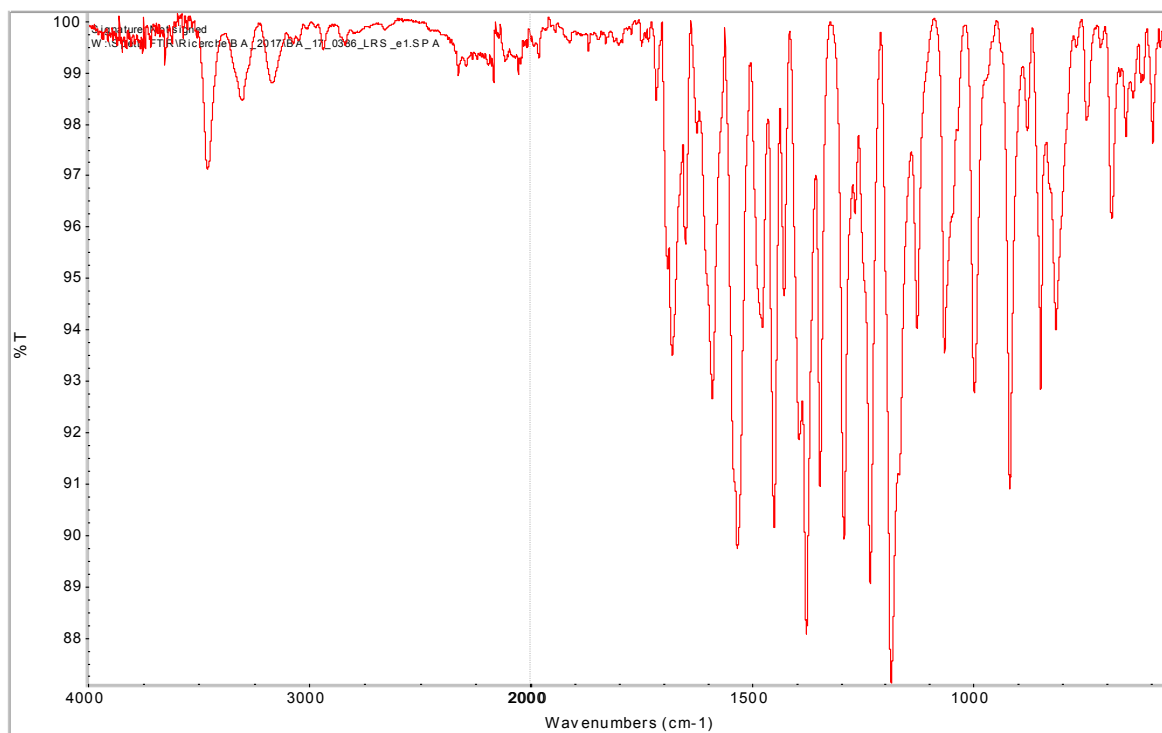
HPLC purity is 95.07%.
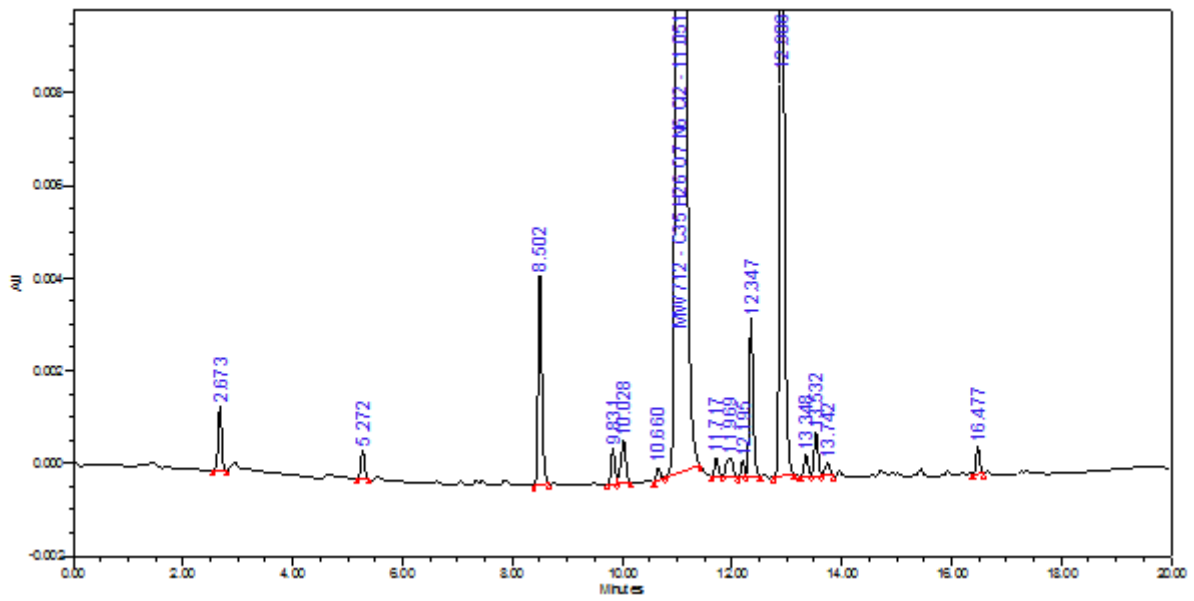
	Name	Int Type	RT (min.)	Area (uV*sec)	% Area
1		bb	2.673	5990	0.194
2		BB	5.272	2895	0.094
3		BB	8.502	21204	0.686
4		BV	9.831	3590	0.116
5		VB	10.028	5548	0.180
6		BV	10.660	1265	0.041
7	MW712 - C35 H26 O7 N6 Cl2	VB	11.051	2914746	94.322
8		BV	11.717	2019	0.065
9		VB	11.969	3397	0.110
10		BV	12.195	1500	0.049
11		VB	12.347	16786	0.543
12		BB	12.906	99904	3.233
13		BV	13.348	2314	0.075
14		VV	13.532	4594	0.149
15		VB	13.742	1743	0.056
16		BB	16.477	2713	0.088

99

References

[1] T. Naito, K. Yoshizawa **2002**, patent US 7683172, patent EP 1683785.

[2] T. Naito, **2003**, patent US 8058474, patent EP 1683785.

[3] ICH harmonized tripartite guideline, Impurities in New drug Products Q3B (R2).

[4] R. Hron, B. S. Jursic, *Tetrahedron Lett.*, **2014**, 55, 1540-1543.

[5] J. Pedersen, *J. Am. Chem. Soc.*, **1967**, 89, 7017-7036.

[6] ICH harmonized tripartite guideline, Impurities: guidelines for residual solvents Q3C (R7).

4. Conclusions

Aim of our project was the introduction of the generic version of Lenvatinib mesylate, an anti-cancer drug used in the treatment of differentiated thyroid carcinoma (DTC), advanced renal cell carcinoma (RCC) and hepatocellular cell carcinoma (HCC). The analysis of the patent situation in the United States revealed the need of a new solid form of the API to anticipate the introduction of the generic version and to apply for a paragraph IV challenge of the solid form patent. Therefore, we discovered a new solid form of Lenvatinib mesylate (ACA-1-HT-dry), which showed high stability at different conditions of temperature and relative humidity. Moreover, from preliminary experiments ACA-1-HT-dry resulted more soluble than the originator's solid forms (Form A and Form C) and hence the new polymorph should not represent an issue for the bioavailability of the API.

At the same time, we developed the preparation process of Lenvatinib mesylate to satisfy the quality requirements of the ICH guidelines (Scheme 4.1).
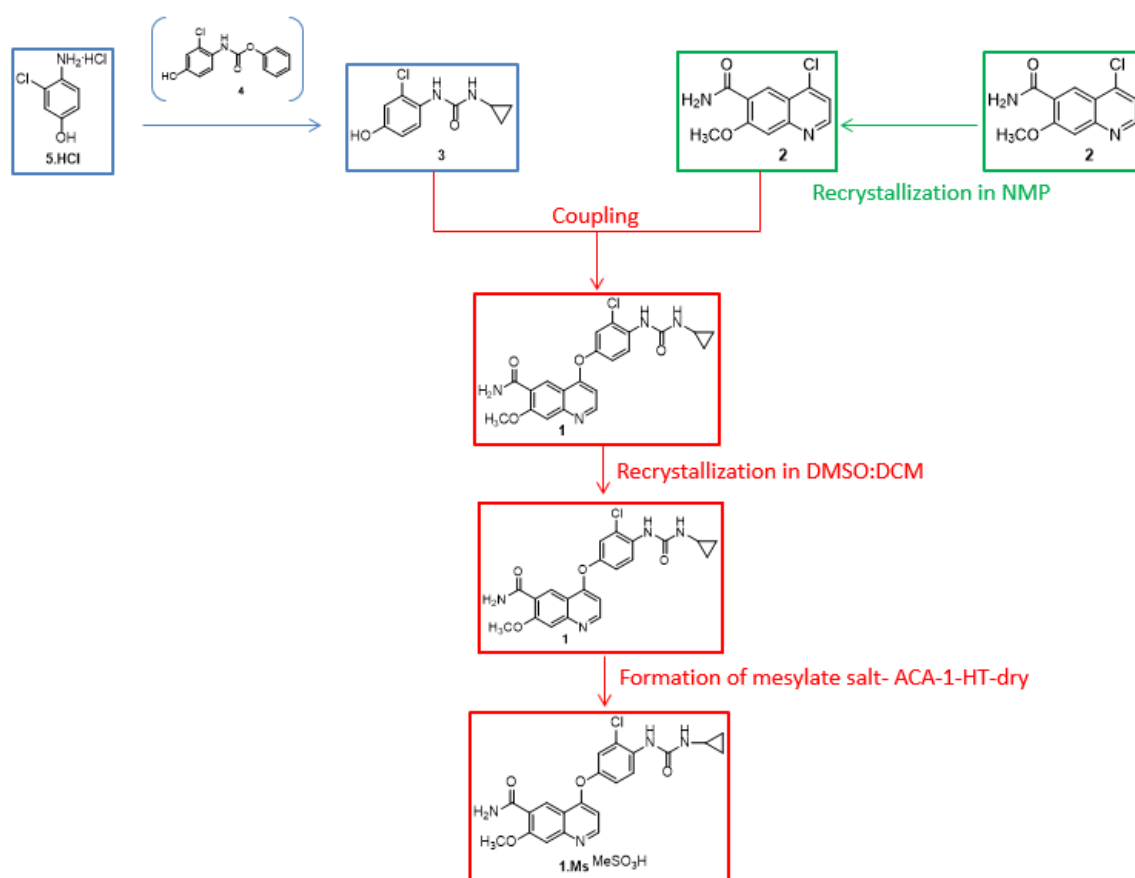
Scheme 4.1. Schematic representation of preparation process of Lenvatinib mesylate ACA-1-HT-dry.

The synthesis of urea **3** from 4-amino-3-chloro-phenol hydrochloride salt **5.HCl** was optimized in order to avoid isolating the intermediate **4** and to achieve a purity higher than 99%. On the other side, the quality of starting 4-chloro-quinoline **2** was standardized by introducing a recrystallization step. Then, Lenvatinib **1** was synthesized by coupling between compound **2** and **3** and this step was critical for the purity of the final API. Therefore, the main impurities were identified, synthesized and characterized and the reaction conditions were modified to minimize their formation. Moreover, a recrystallization step was introduced to further purify Lenvatinib **1** before the preparation of the mesylate salt **1.Ms** in the new ACA-1-HT-dry form. The optimized process allowed to achieve the required purity and was scaled-up in high containment Kilo-Lab to obtain 1 Kg of Lenvatinib mesylate ACA-1-HT-dry form.

In conclusion, we developed a scalable process for the preparation of a generic version of Lenvatinib mesylate, which is characterized by the required purity and by a new, not infringing and stable solid form.

SECTION B: SULFATED POLYSACCHARIDE (SP)

1. Introduction

The second topic of the thesis concerns a sulfated polysaccharide (from now on referred to as "SP"), mainly made up of sulfated pentose sugars, predominantly xylose, (a generic structure is reported in Figure 1.1). The polymer may have anticoagulant, fibrinolytic and anti-inflammatory properties.
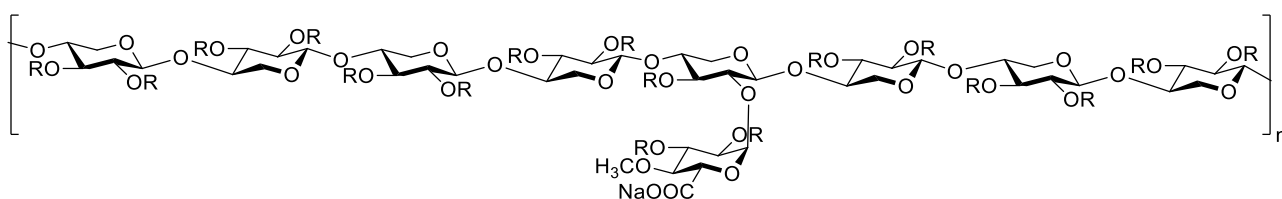
Figure 1.1. Structure of the sulfated polysaccharide. R = SO_3Na, Ac, H.

1.1 Aim of the project

The aim of this part of the work is the replication of a brand name SP (from now on referred to as the "Originator" one) commercialized in highly regulated markets (i.e. US and EU).

Recently the US Food and Drug Administration has become aware of the low number of generics present on the market when the API is not a well-defined molecule. Cases where the APIs are complex entities (e.g. peptides, iron complexes, natural derived compounds and polymers) prompted FDA to issue documents and guidelines to encourage API generic producers to enter these fields.[1,2]

SP, being a mixture of polymers, clearly belongs to this last class of compounds.

According to FDA, for these products the equivalence between brand name API and the generic one must reflect three main aspects:

➢ *The source of the starting material*. The sulfated polysaccharide is obtained by a semi-synthetic process. The biomass from which the polysaccharide is extracted has to be the same for the generic and originator products.

➢ *Physicochemical properties*. The overall structural properties and fingerprints should be comparable.

> *Building blocks composition and chain linkage*. The two products should contain the same monosaccharide units (e.g. xylose, 4-*O*-methylglucuronic acid) linked together in a comparable way (e.g. branched units, anomeric configuration).

To study the last two points, several orthogonal analytical methods are necessary. These will be discussed in chapter 2. They are crucial to understand the key structural features of the brand product, and also to demonstrate the sameness of our generic version compared to the originator one. The process will be discussed in the third chapter and consists in the isolation of the glucuronoxylan from the biomass, followed by sulfation and purification steps.

1.2 Pharmacological properties

SP is chemically similar to glycosaminoglycans (GAGs) and hence able to mimic several of their pharmacological activities such as the anti-coagulant and fibrinolytic actions. Specifically, the anticoagulant activity is similar to that of Dextran sulfate, yet it is 10-15 times less active than the benchmark molecule heparin (Table 1.1).[3]

Compound	Anticoagulant activity (U/mg)
Sulfated polysaccharide (SP)	14.4
Dextran sulfate (MW 5000)	14.7
Dextran (MW 90000)	0.01
Heparin (MW 11000)	177
N-desulfated heparin (MW 8800)	0.6

Table 1.1. Anticoagulant activity of a few polysaccharides

SP exert its anti-coagulant activity through both Antithrombin III-dependent and Antithrombin III-independent mechanisms (Figure 1.2). Antithrombin III (AT III) is a glycoprotein present in the plasma which is able to inhibit sequentially coagulation factors IXa, Xa, XIa and XIIa and eventually thrombin. The latter transforms fibrinogen into fibrin monomers which polymerize generating insoluble clots. Thanks to this cascade of reactions, compounds that, like heparins, can activate antithrombin III exhibit anti-coagulant activity. SP is also able to activate AT III, increasing thrombin inhibition by 40-folds.[4] This value is very low compared to the 50,000-fold increase generated by heparin[5] and indicates the presence only of a weak interaction between SP and AT III.

However, SP acts also through an antithrombin III- independent mechanism, by interacting with Heparin Cofactor II (HCII). This cofactor is a serine protease inhibitor which inhibits rapidly thrombin when activated. The electrostatic affinity of HC II with SP is greater than with heparin and dermatan sulphate and hence SP can activate HCII at lower concentration.[6]

Figure 1.2. Control mechanism of coagulation. The sulfated polysaccharide can potentiate the activity of both Antithrombin III (AT) and Heparin Cofactor II (HCII).

Another interesting pharmacological property of SP is still linked to the structural similarity with glycosaminoglycans (GAGs), which play a key role for urothelium, the epithelial tissue which lines the organs of the urinary system. Indeed, bladder epithelium is protected by a GAG layer which controls cell permeability, avoiding the contact of hazardous solutes (e.g. toxins) with cell membranes. A damage to this barrier may cause a higher permeability to urothelium leading to severe pathologies, like Interstitial cystitis (IC) (Figure 1.3). An exogenous sulfated polysaccharide can repair the damage by occupying the spaces in the GAGs layer and hence regenerating the barrier to bladder epithelium.[7]

Figure 1.3. Mechanism of urothelium inflammation. (Figure modified from Reference 8)

Moreover, SP decreases the secretion of histamine by mast cells, thus reducing the inflammation associated with interstitial cystitis. [9]

Anti-inflammatory properties can also be exploited in other pathologies such as rheumatoid arthritis where SP inhibits the nuclear factor κB; a factor that mediates the inflammation in chronic diseases.[10] Furthermore, in osteoarthritis SP is also effective in cartilage protection inhibiting the degradation of aggrecan, a proteoglycan which is the main component of an articular cartilage.[1]

SP showed also promising results in the treatment of allergic rhinitis, an inflammation of the upper airways.[11] Indeed, in vitro studies demonstrated the ability of SP in binding Th2 cytokines, a crucial mediator in allergic inflammations. In vivo studies have shown anti-inflammatory effects similar to the corticosteroid budenoside, which is commonly used in the treatment of this disease.

Finally, a few studies showed potential anti-neoplastic activity.[12] Angiogenesis, the formation of new blood vessels, is fundamental for the growth of neoplastic tissues and is stimulated by heparin-binding growth factors (HBGFs). SP binds HBGFs, inhibiting their activities, blocking angiogenesis and not allowing the growth of tumor cells. Thanks to this mechanism SP is effective in preventing tumor formation in animals inoculated with moderate numbers of tumorigenic cancer cells. However, SP showed low effect in the presence of high number of tumorigenic cancer cells or with well-established tumors probably because the cells express a number of HBGFs too high to be inhibited effectively. [13]

As far as the pharmacokinetic is concerned, SP has a low bioavailability (0.5-1% at steady state for an oral administration of 400 mg daily).[14] In humans it is mainly distributed in the genitourinary tract, bone marrow, lungs, liver, periosteum, skin and spleen. Partial desulfation of the drug has been observed (68% of the dose) in the liver and spleen.[8]

1.3 Patent and market situation

The Originator developed the API in 1947 and since then has kept the structure of the drug and the preparation method as a trade secret so that, no patents have been filed (patent filing would have required disclosure of the structure and/or the preparation method). Nowadays, other companies or institutions have filed patents relating to the use of sulfated polysaccharides, possible preparation processes or part of them (for example patent appl. WO2016/184887[15] or WO2018/043667[16]). However, as far as we know, none of them was able to replicate the Originator SP structure. In the US the Originator SP received the orphan drug status for the treatment of interstitial cystitis. The seven years orphan drug exclusivity period expired in 2007.

Therefore, since that date no one has been able to introduce in highly regulated markets a SP generic version. This is probably due to two main reasons: the first one is associated to the demonstration of sameness (a requirement for a generic API), which means that the API structure of the generic version must be exactly the same as the Originator one. In this case the difficulty stems from the fact that the API is not a well-defined chemical entity, with a precise a structure, but a dispersion of polymeric chains. The second obstacle is the demonstration of bioequivalence because, to date, there are no published methods to quantify the API in biological fluids. Performing a clinical trial is the recommendation of the FDA to solve this issue and clearly demonstrate the effectiveness of the generic drug.[17]

United States (3-4 tons/year) and Europe (about 1 ton/year) are the main markets for the Originator and also where there is not generic competition. On the other hand, SP generic versions are present in Latin America and India. However, the quality and the structural features of these products are not matching the Originator ones. Indeed, samples of the SP coming from US, EU, Latin America and India were compared by ^{1}H-NMR (Figure1.4) and capillary zone electrophoresis (CZE).
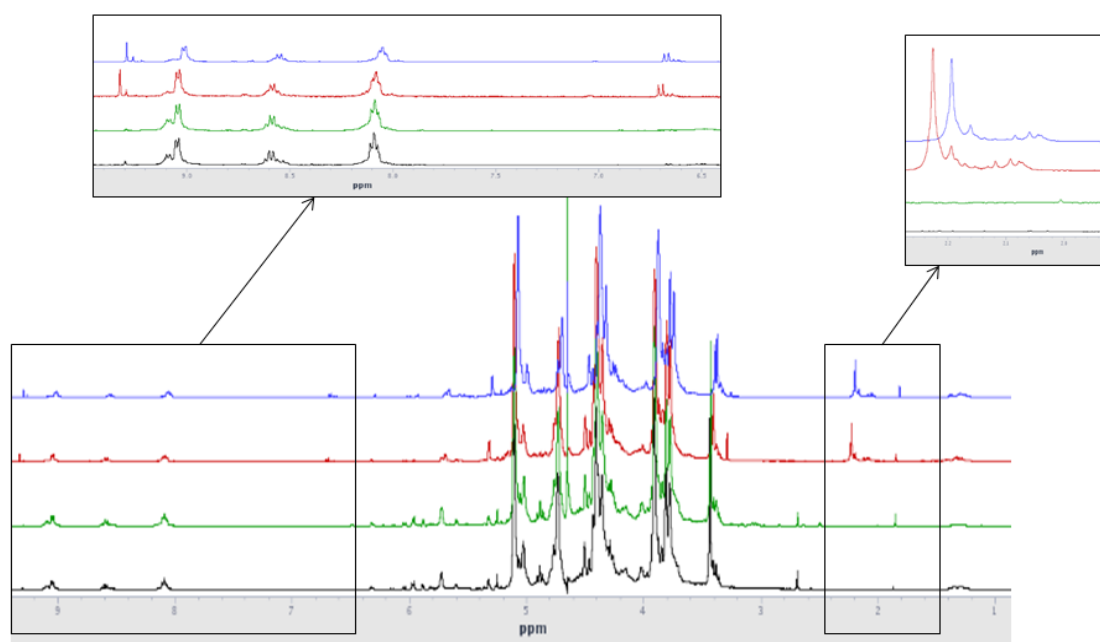
Figure 1.4. 1H-NMR overlay of sulfated polysaccharide coming from US (blue), EU (red), Latin America (green), India (black). In the two squares there are the zoom of the aromatic and double bonds region on the left and of acetyl region on the right.

The proton spectra show a clear difference at about 2 ppm, where there is a group of signals in the US (blue) and EU (red) related to presence of acetyl moieties, while there are not signals in the

Latin American (green) and Indian (black) products. Moreover, the proton spectra show other minor differences: for example, only in the US and EU API there are two groups of signals at 9 ppm and 6.7 ppm due to the presence of aldehydes and double bonds respectively. Another technique considered crucial for the characterization of SP is capillary zone electrophoresis (CZE) (this technique will be discussed in Chapter 2.4). The latter confirmed clearly the differences between the APIs (Figure 1.5). In the products coming from United States and Europe there is a well-defined pattern of peaks, while in the ones from Latin America and India the electropherogram is characterized by peaks with low resolution and intensities.
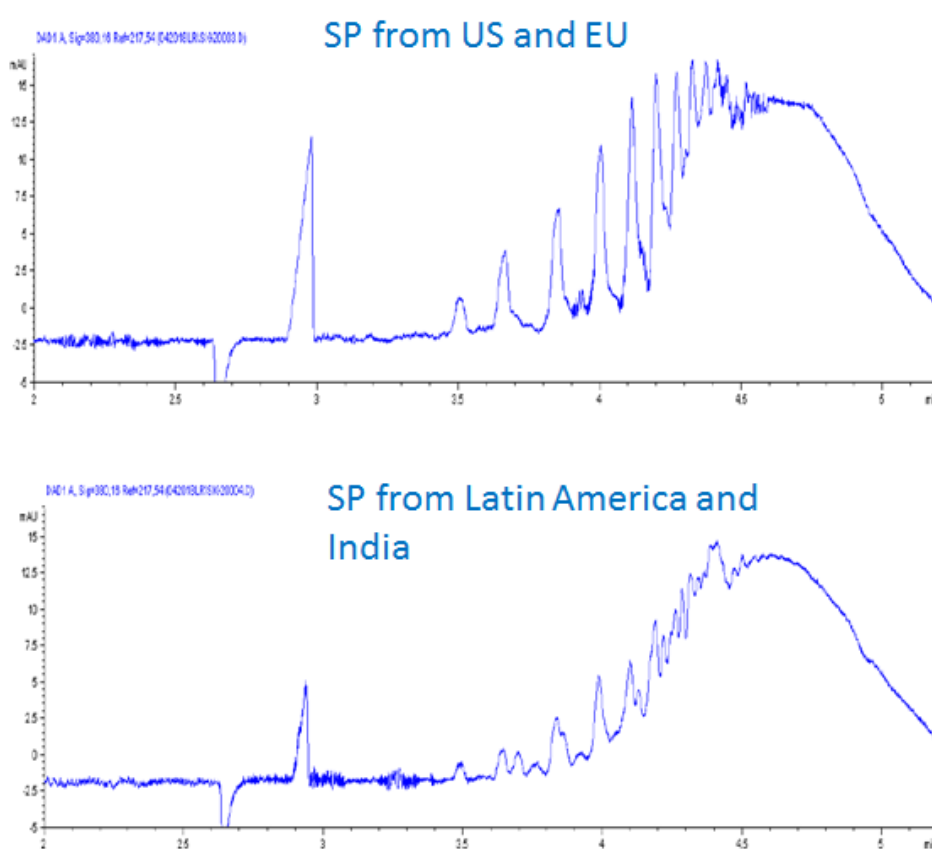
Figure 1.5. CZE comparisons of sulfated polysaccharides coming from different countries

Clearly the APIs present in Latin America and India are substandard products, which do not need to comply to the quality requirement of United States and Europe.[18]

References

[1] X Jiang, Introduction to Complex Products and FDA considerations, *Oral communication at "Demonstrating Equivalence of Generic Complex Drug Substances and Formulations" meeting*, October 6th, **2017**.

[2] D. Zhang, Demonstrating complex API sameness, *Oral communication at "Demonstrating Equivalence of Generic Complex Drug Substances and Formulations" meeting*, October 6th, **2017**.

[3] R. Parthasarathi, A. Jayakrishnan, *Biodegradable polymers in Clinical Use and Clinical Development*, Wiley, **2011**, 185-216.

[4] M. Scully, V. Kakkar, *Biochem. J.*, **1984**, 571-578.

[5] M. Hoylaerts, W. Owen, D. Collen, *J. Biol. Chem.*, **1984**, 259, 5670-5677.

[6] M. Scully, V. Kakkar, *Thromb. Res.,* **1986**, 41, 489-499.

[7] C. Parsons, *Urol. Clin. North Am.*, **1994**, 21, 93-100.

[8] V. Anderson, C. Perry, *Drugs*, **2006**, 66, 821-835.

[9] G. Chiang, P. Patra, R. Letourneau, *Adv. Exp. Med. Biol.*, **2003**, 539, 713-729.

[10] P. Barnes, M. Karin, *N. Engl. J. Med.*, **1997**, 336, 1066-1071.

[11] C. Sanden, M. Mori, P. Jogdand, J. Jonsson, R. Krishnan, X. Wang, J. Erjefalt, *Immunity, Inflammation and Disease*, **2017**, 300-309.

[12] W. Figg, J. Pluda, O. Sartor, *Antiangiogenic Agents in Cancer Therapy*, Humana Press, **1999**, 371-383.

[13] S. Swain, A. Wellstein, B. Parker, M. Lippman, C. Steakley, R. deLap, *Ann. N. Y. Acad. Sci.*, **1993**, 63-70.

[14] M. Simon, R. McClanahan, J. Shah, *Xenobiotica*, **2005**, 35, 775-784.

[15] L. De Ferra, E. Ammirati, S. Andreassi, M. Annibaldi, L. Mandelli, B. Pinto, F. Stracqualursi, **2015**, patent appl. WO2016/184887.

[16] T. Ishikawa, T. Kashiwamura, T. Kato, T. Koga, O. Ishikawa, **2016**, patent appl. WO2018/043667.

[17] Draft Guidance, Food and drug administration, September **2012**.

[18] J. Fareed, H. Nader, E Coyine, *Clin. Appl. Thromb. Hemost.*, **2005**, 11, 363-366.

2.Investigation of API structure

The sulfated polysaccharide (SP) **1** is a polymer characterized by a main chain of xylose units linked by β (1→4) glycosidic bonds. There are ramifications of 4-*O*-methylglucuronic acid (MGA), which depart from position 2 of the xylose units (Figure 2.1). The chains are highly, but not completely, sulfated with a few acetylated hydroxy groups. Figure 2.1 shows the most represented structure, but there are other different saccharide units or substituents intercalating the main chain.
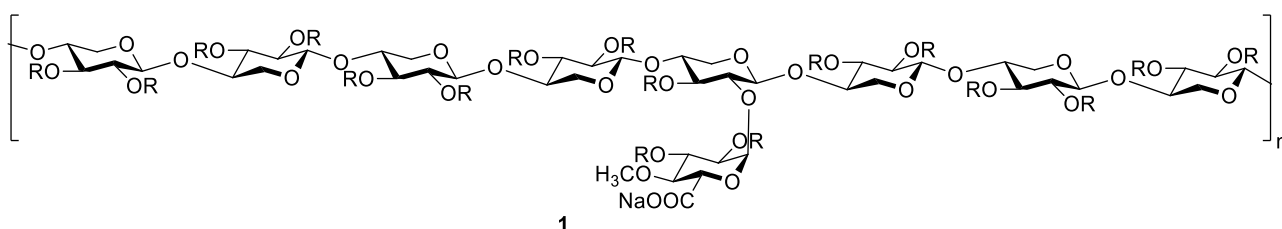
Figure 2.1. Structure of sulfated polysaccharide (SP), R = SO_3Na or H or Ac.

As already anticipated, a deep understanding of the structure is necessary to develop the project and requires orthogonal analytical methods. Therefore, we recovered several batches of the originator's product commercialized in the United States and we extracted the API from the capsules. Each capsule contains 100 mg of API and about 130 mg of excipients (microcrystalline cellulose and Magnesium stearate). The sulfated polysaccharide is highly soluble in water while the excipients are insoluble in aqueous media. Therefore, API can be obtained from the capsules suspending their content in water at r.t. for 20h, filtering off the insoluble residue and freeze-drying the filtrate. This procedure was applied to ten different batches of the commercial product and the recovery of API was 95-99%.

In order to fully elucidate the structure, the extracted SP was subjected to Gel Permeation chromatography for Molecular weight determination, Nuclear magnetic Resonance (NMR), determination of Sodium and Sulfur content, capillary zone electrophoresis (CZE), liquid chromatography-mass spectrometry (LC-MS).

This work was conducted in collaboration with Ronzoni institute, especially for NMR studies, and the analytical group of Indena, who developed the analytical methods and elaborated Principal

Component Analysis. My main contribute was in the interpretation and overall evaluation of the data.

Each technique will be discussed in detail in the next paragraphs, but the exact conditions of the analytical methods will not be presented for confidentiality reasons.

2.1 Gel Permeation Chromatography (GPC)

SP is not a single chemical entity with a defined chain length, but it is a dispersion of polymeric chains with unequal lengths and hence with different molecular weight. The distribution of molecular weight can be described by two main parameters. The first one is the Number Average molecular weight (Mn) which can be defined by:

$$Mn = \frac{\sum NiMi}{\sum Ni}$$

where Mi is the molecular weight of a chain and Ni is the number of the chains with that molecular weight. This value represents the average molecular weight of all the chains of a polymer.

The second key parameter is the Weight average molecular weight (Mw) which can be defined by:

$$Mw = \frac{\sum NiMi^2}{\sum NiMi}$$

In contrast to Mn, Mw considers the molecular weight of a chain in the determination of the contribute of a chain to the average molecular weight. This means that a chain with a higher molecular weight contributes to Mw more than a chain with a lower molecular weight. For all polydisperse polymers Mw is a higher value than Mn.

The ratio between Mw and Mn is defined as polydispersity index (PD), which measures the broadness of molecular weight distribution. A high polydispersity means a large distribution with a lot of chains with different lengths, while PD equal to 1 indicates a monodisperse polymer with chains of the same length. Another useful parameter is molecular weight of the highest peak (Mp) which is the mode of the molecular weight distribution.

All these parameters can be determined by Gel Permeation chromatography (GPC) or Size Exclusion Chromatography (SEC), a type of liquid chromatography based on porous structures used as stationary phase, which separate the chains according to the dimensions (Figure 2.2).[1] The polymer is dissolved in the mobile phase and meets pores of different sizes in the stationary

phase. The biggest chains cannot go into the pores and are just carried to the end of the column by mobile phase eluting first. On the other hand, the smallest chains enter in a few pores and hence are slowed down in their run.
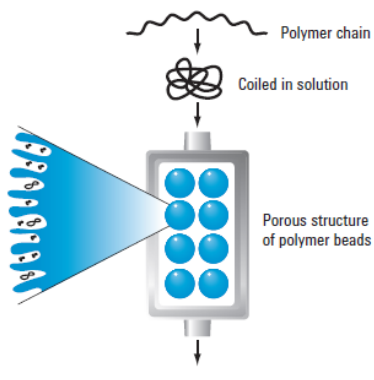
Figure 2.2. Principle of polymer separation in GPC. Figure modified from Reference 1

The eluted chains can be detected in different ways and, specifically, two different detectors were used for SP. The first one is a refractive index (RI), which is based on changes in the refractive index of the column eluate. It is a universal detector because is capable to detect any kind of molecules, but it requires calibration to return the values of MW distribution. Pullulans, polysaccharides of maltotriose, with different molecular weights (708 KDa, 200 KDa, 21 KDa, 10 KDa, 6 KDa) can be used as standards to generate a correlation curve between elution time and molecular weight of the polymers. This approach is based on the assumption that GPC separate polymer chains according to their hydrodynamic volume which in turn correlates with the molecular weight. However, we know that the chemical structure of the polymer can influence the conformation of the chains modifying the elution time. For example, with an ionic polymer like SP, due to different conformations, 10 KDa chains can have the same hydrodynamic volume and hence the same elution time of 5 KDa pullulans. Therefore, the MW values obtained with RI are not absolute and strongly depend on chromatographic conditions. This detector is anyway useful for a relative comparison between the Originator SP and the SP prepared during lab trials, because the chemical properties of the two polymers are almost identical.

The second detector used for MW determination is Multi Angle Laser Light Scattering (MALLS).[2] This detector uses a laser beam, which irradiates the eluate, and devices able to collect the scattered light at different angles. The scattered light is evaluated as Raleigh ratio (RΘ), the excess scattering of the sample and solvent compared to solvent alone. Raleigh ratio is directly proportional to molecular weight, a constant K* and the concentration.

$$R\Theta = Mw\, K^* c \ \ where\ K^* = (dn/dc)^2 K$$

The refractive index increment (dn/dc) is the degree by which the refractive index of a solution varies as the concentration of the solution changes. This parameter is considered constant for a polymer and solvent combination and can be experimentally determined with few injections at different concentrations.[3] Once the refractive index increment is known, it is possible to determine the molecular weight of the sample from the scattered light (RӨ).

As mentioned before, both detectors were used on originator's sulfated polysaccharide and the chromatograms are shown in Figure 2.3. Moreover, the results obtained for ten different batches and in different analysis sessions are reported in Table 2.1.
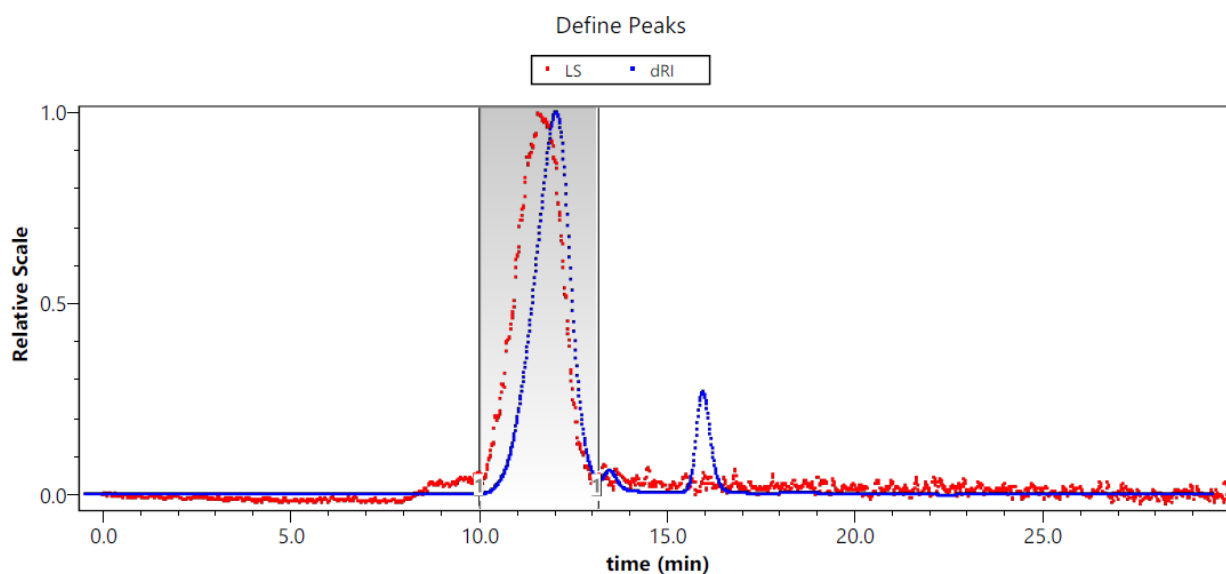
Figure 2.3. GPC chromatograms of commercial sulfated polysaccharide. In blue the refractive index (RI) chromatogram and in red the light scattering one (MALLS). The peak of the product is in the grey area.

Technique	Mn (KDa)	Mw (KDa)	Mp (KDa)	PD
GPC-RI	3.8 - 5.7	7.8 - 8.4	5.1 - 5.9	1.5-2.0
GPC-MALLS	3.7 - 3.9	5.5 - 5.9	-	1.5

Table 2.1. GPC results of commercial sulfated polysaccharide with refractive index (RI) and light scattering (MALLS).

In both RI (blue) and MALLS (red) chromatograms it is possible to observe a main peak at about 10 minutes, but in RI two additional peaks with lower intensity related to salts (sulfates and acetates) are present. MALLS is more sensitive to higher molecular weights and hence it does not detect compound with very low MW, like inorganic salts.

The values obtained with refractive index detector show a higher variability than the ones obtained using a light scattering detector. For example, Mn according to RI detector can be a value between 3.8 and 5.7 KDa (Table 2.1), while Mn using a MALLS detector is between 3.7 and 3.9

KDa. This behavior confirms that the refractive index detector, returning MW values relative to the pullulans calibration, is very sensitive to elution conditions and column age so that there is not high reproducibility among different analytical sessions. On the other hand, MALLS yields more reproducible, more accurate and absolute results. Therefore, we can conclude that the commercial sulfated polysaccharide has a molecular weight between 3.7 and 5.9 KDa and a polydispersity equal to 1.5.

GPC analysis can also give additional information if a viscosimeter is used to determine intrinsic viscosity of the polymer. This parameter allows to build Mark-Houwink-Sakurada plot (Figure 2.4) which correlates intrinsic viscosity with molecular weight.
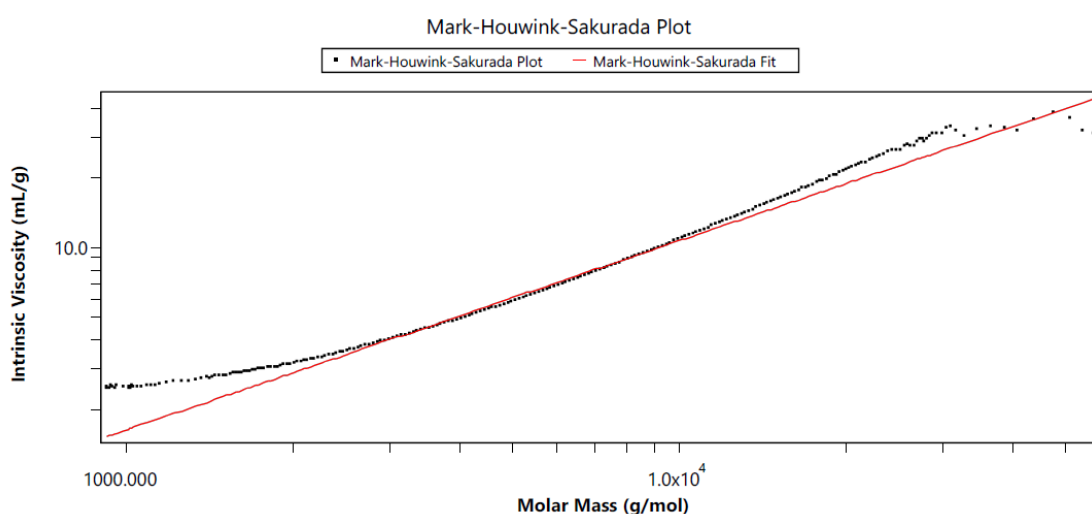
Figure 2.4. Mark-Houwink-Sakurada plot of branded sulfated polysaccharide.

This plot allows to determine two parameters (α and K) which describe the behavior of a polymer in solution. The sulfated polysaccharide has α equal to 0.82, which is a typical value of a random coil in solution. A higher α value (2.0) is obtained for "rigid rod", while a lower α (<0.5) is associated to compact sphere.

2.2 Nuclear Magnetic Resonance (NMR)

A key technique for structural characterization of the sulfated polysaccharide is Nuclear Magnetic Resonance (NMR). Useful data were already present in patent application WO 2014/114723[4] but further work needed to be done aiming at:

a) Identification of the signals in proton spectrum and hence of the species present in SP.

b) Relative quantification of key features through integral ratios

c) Study of the consistency of commercial batches

d) Identification of monosaccharides compositions

NMR characterization of the product was carried out in collaboration with Ronzoni institute – Institute for Chemical and Biochemical Research.

2.2a Identification of ^{1}H-NMR signals

The proton spectrum of the Originator SP was recorded in D_2O and is shown in Figure 2.5. It can be divided in 4 main regions: the first one, between 10 and 6.5 ppm (blue) includes aldehydes, aromatic signals and double bonds (minor features); in the second one, between 6.5 and 5.0 ppm (green) there are the anomeric protons of the glycosidic units; in the red area, between 5.0 and 3.5 ppm there are the other protons of sugars and finally, between 3.5 and 1.0 ppm (yellow) there are signals related to the methyl groups of 4-O-methylglucuronic acid and acetyl moieties.
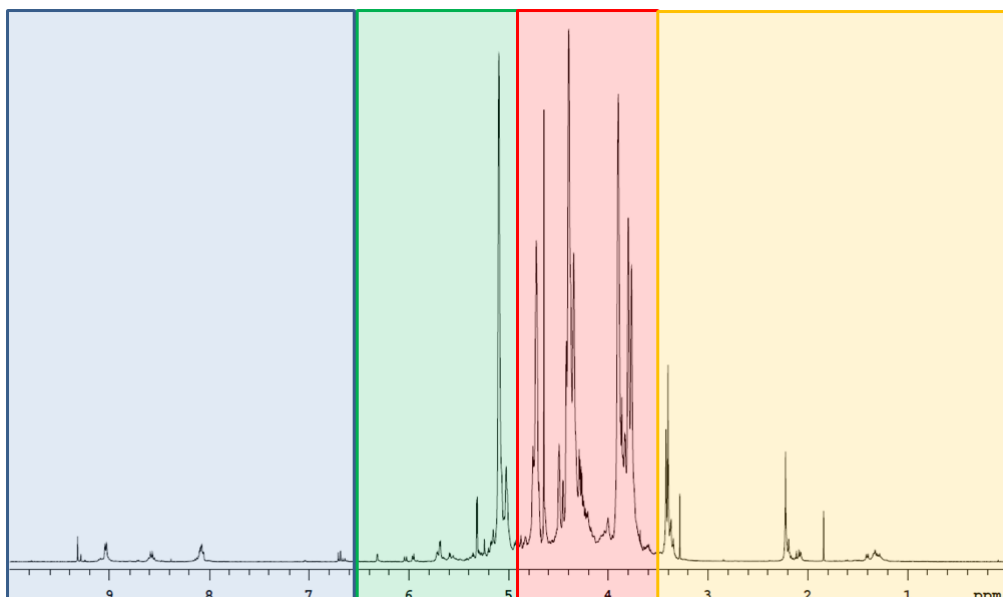
Figure 2.5. ^{1}H-NMR spectrum of the sulfated polysaccharide. There are four main areas: aldehyde, aromatic and double bonds (blue), anomeric protons (green), other sugar protons (red), methyl groups (yellow).

In the blue area there is a singlet at 9.38 ppm and this chemical shift suggests the presence of an aldehyde (Figure 2.6). The nature of this species was further clarified by heteronuclear 2D NMR (Figure 2.7) because in HSQC (Heteronuclear Single Quantum Correlation) it is observable the correlation between the proton at 9.38 ppm and the carbon at 193 ppm and in HMBC (Heteronuclear Multiple Bond Correlation) the carbon at 193 ppm is correlated to the doublet at 6.8 pm, the protons of a double bond. Therefore, we can assume that there is a double bond conjugated to an aldehyde and hence species with structure **2** (Figure 2.6)**,** an open ring aldehyde with C2-C3 double bond, are very plausible. The second cluster of species present in SP, which is evident in this region are related to three signals between 9 and 8 ppm, which belong to the

protons of a pyridinium ring attached to the terminal anomeric position of xylose **3**. Indeed, sulfation of the polysaccharide is commonly performed in pyridine and, as reported in the literature[5], substitution of hydroxyl groups at the anomeric position with pyridinium moieties occurs.
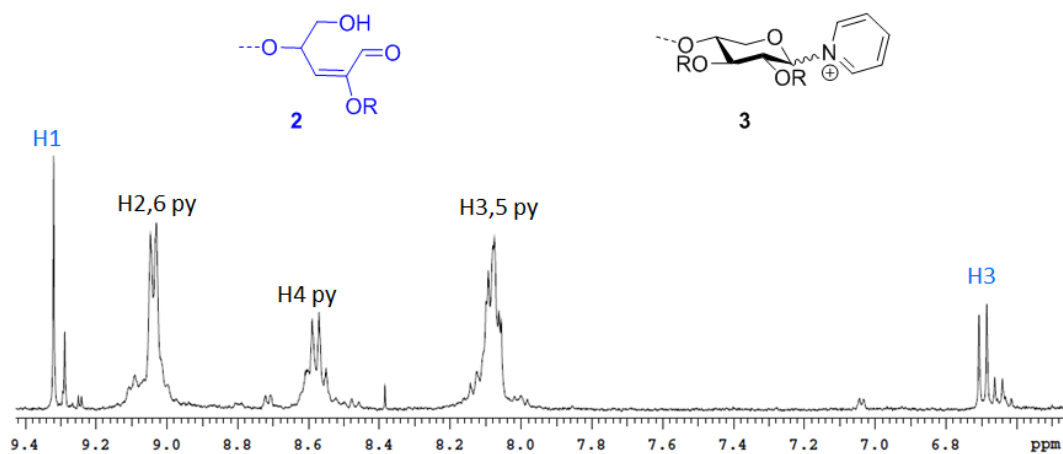
Figure 2.6. Identification of signals in ^{1}H-NMR spectrum (9.4-6.7 ppm). The signals are related to presence of an open ring aldehyde with C2-C3 double bond (blue) and pyridinium species (black). R =SO$_3$Na
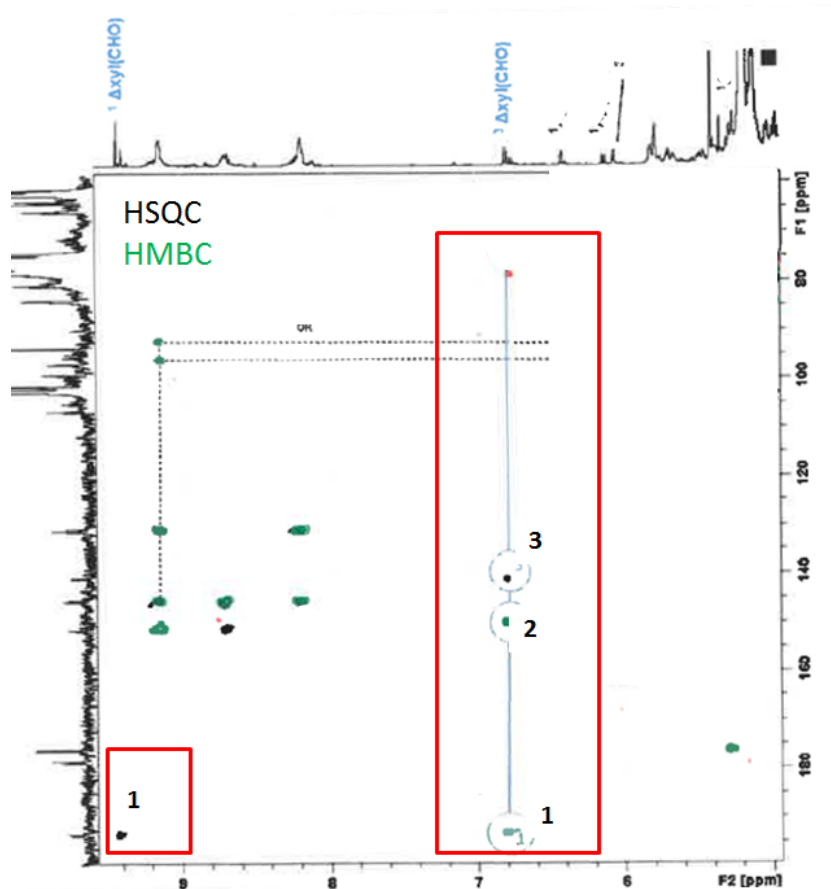
Figure 2.7. HSQC and HMBC of SP (9.4-6.7 ppm). In HSQC it is observable the correlation between the proton at 9.38 ppm and the carbon at 193 ppm and in HMBC the carbon at 193 ppm is correlated to the doublet at 6.8 pm.

In the green region, between 6.5 and 5.0 ppm the anomeric protons of the main glycosidic units (Figure 2.8) are found. As described in the literature[4], the most intense signals are related to anomeric protons of xylose units sulfated in position 2 and 3 and located at the not reducing end or in the internal portion of the chain. At higher fields there is a group of signals identified as xylose anomeric protons at the reducing end in β configuration where xylose is sulfated at position 2 and acetylated at position 3. It is worth noting that acetylation is a key feature of the Originator SP and the acetyl groups are mainly located at position 3 of xylose units (this aspect will be better explained in the discussion of the yellow portion of the spectrum). At 5.3 ppm there is a doublet due to the presence of α anomers of sulfated xylose.
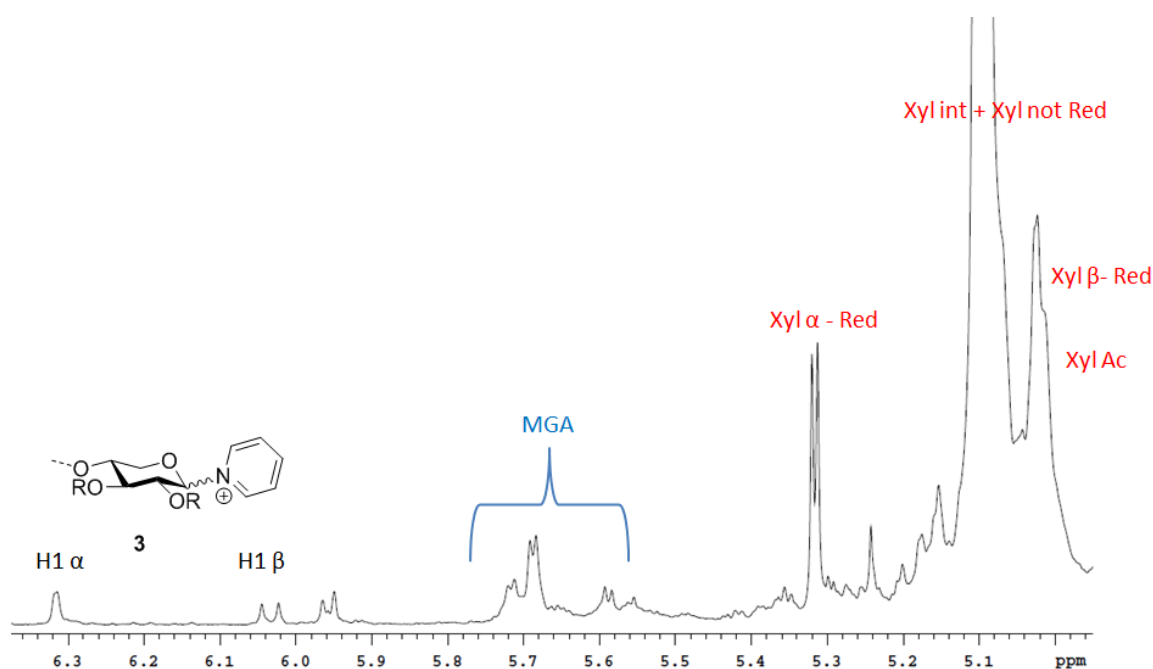
Figure 2.8 Anomeric region of ^{1}H-NMR spectrum. MGA = 4-O-methylglucuronic acid; Xyl α or β-Red = sulfated reducing terminal xylose as α or β anomer; Xyl int = internal sulfated xylose; Xyl Ac = acetylated xylose.

The signals between 5.8 and 5.5 ppm belong to the anomeric protons of 4-*O*-methylglucuronic acid (MGA) attached to position 2 of xylose. The anomeric proton of MGA does not give a single signal, but three main signals depending on which kind of xylose they are linked to (e.g. sulfated in positions 2 and 3 or sulfated in 2 and acetylated in 3). Finally, the signals at 6.3 and 6.0 ppm belong to the anomeric protons of structure **3**. HMBC shows correlation between the ortho protons of the pyridine ring and the carbons at anomeric reducing end units confirming the presence of a covalent bond between the pyridinium moiety and the sugar (Figure 2.9).
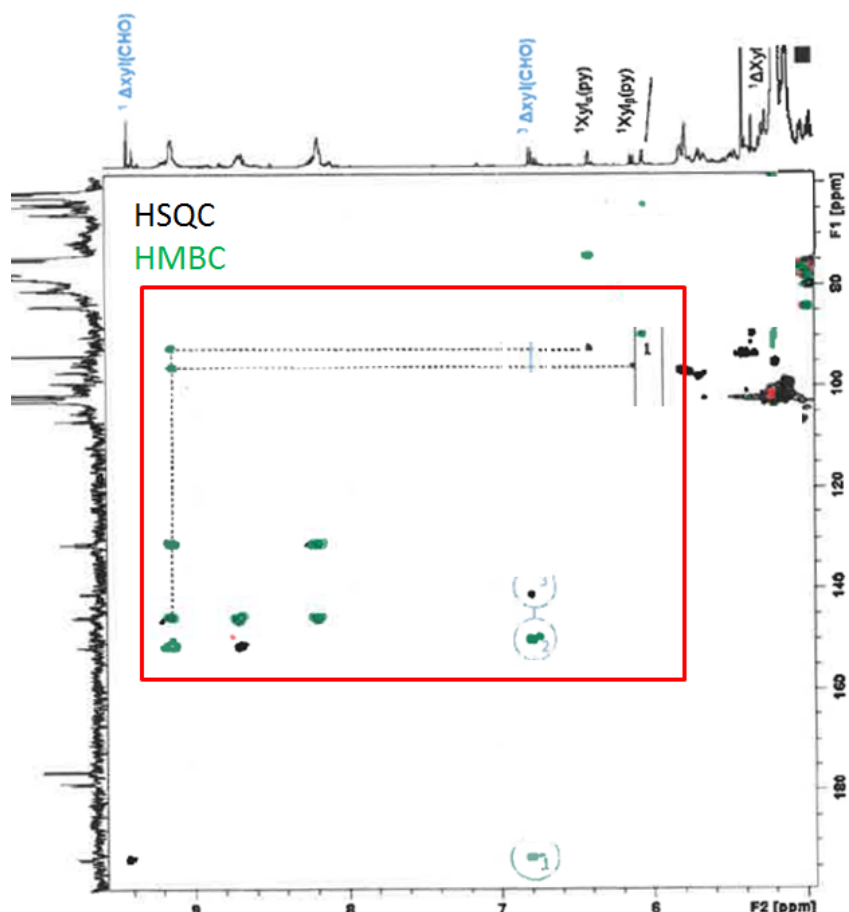
Figure 2.9. HSQC and HMBC of SP (9.4-6.7 ppm) HMBC shows correlation between the ortho protons of pyridine ring and the anomeric carbons at 100 ppm. These carbons are correlated in HSQC to protons at 6.3 and 6.0 ppm.

In the 5.0 – 3.5 ppm region of the spectrum (red) there are the protons signals in the other positions of the xylose ring (H2- H5). For each position there are signals of sulfated internal xylose, sulfated reducing end xylose, sulfated xylose attached to MGA, acetylated xylose and 4-*O*-methylglucuronic acid.

Finally, between 3.5 and 1.0 ppm (yellow) there are the signals of various methyl groups (Figure 2.10). Around 3.5 ppm there are the signals of CH_3 of methoxy groups of MGA. It is not a simple singlet, but two singlets with less intense signals at lower ppm. Indeed, as seen for the anomeric protons, methyl signals are also split according to the position of 4-*O*-methylglucuronic acid in the chain and in particular the major splitting is related to the main chain acetylation. The signal at lower field is assigned to methyl groups of MGA attached to not acetylated xylose and the other one is methyl of MGA attached to acetylated xylose.
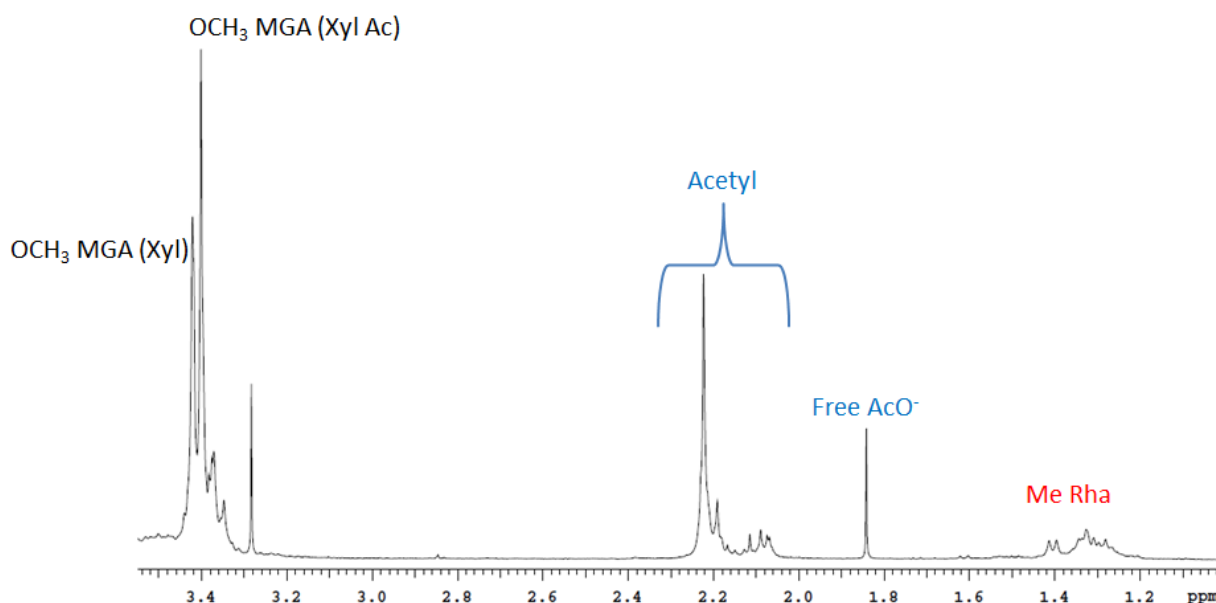
Figure 2.10. Methyl groups region of ^{1}H-NMR spectrum. MGA (Xyl) = 4-O-methylglucuronic acid attached to xylose; MGA (Xyl Ac) = 4-O-methylglucuronic acid attached to acetylated xylose; Rha = rhamnose

The signals of acetyl groups are present at about 2.2 ppm. The most intense singlet are the acetyl groups present at position 3 of a xylose unit which bears 4-*O*-Methylglucuronic acid in position 2. In fact, HMBC experiment shows correlation between both the position 3 of the acetylated xylose units and a carboxyl group and between position 2 of the latter xylose units and glucuronic acid residues. This evidence was crucial to identify the main location of acetyl groups and to explain the splitting of methyl signals of 4-*O*-methylglucuronic acid. The other signals at 2 ppm are acetyl groups probably present at position 3 of xylose units located in different points of the chain. The singlet at 1.85 ppm is related to free acetates. ^{1}H-NMR spectrum of SP spiked with Sodium acetate confirms the assignment.

The last group of signals at 1.3 ppm are related to the 6-methyl of Rhamnose units. Indeed, the sulfated polysaccharide is mainly characterized by xylose unit, but also contains other carbohydrates which were investigated by NMR after hydrolysis (paragraph 2.2.d).

2.2b Relative quantification by ^{1}H-NMR

NMR also allows to quantify the most important features of the product. The following are measured: 3-*O*-acetylation of xylose bearing MGA, overall degree of acetylation, amount of 4-*O*-methylglucuronic acid, average degree of polymerization and amount of pyridinium species.

- 3-*O*-acetylation can be calculated in two different ways. The first one uses the singlet at 2.2 ppm. The integral of this signal is divided by the sum of integrals of methyl groups of MGA

attached to xylose both acetylated and not acetylated (two singlets at 3.4 ppm). The second method of calculations is exclusively based on signals of methyl groups of MGA: (OCH₃ MGA (xyl Ac) *vs* sum of OCH₃ MGA (xyl Ac) and OCH₃ MGA (Xyl)).

- Overall degree of acetylation can be calculated as the ratio between the integrals of all acetyl groups (2.3-2.0 ppm) and the sum of the anomeric protons of xylose (internal and not reducing xylose, reducing xylose and acetylated xylose). However, the signals of acetylated xylose and β anomer of xylose at the reducing ends are overlapped with other unknown signals and hence need to be calculated indirectly. Acetylated xyloses are determined using the integral of acetyl at 2.2 pm, while β anomers are calculated from the integrals of α anomers assuming the α/β ratio equal to 35/65, the value of the xylose monosaccharide.

- Methylglucuronic acid can be quantified using the integrals of the anomeric signals at 5.75-5.60 ppm or of the methyl groups at 3.4 ppm versus the anomeric protons of xylose, calculated as previously described.

- The average degree of polymerization can be calculated as the ratio between the integrals of anomeric protons of xylose reducing ends and the sum of the anomeric protons of all xylose units.

- The amount of pyridinium species **3** can be calculated from the integral of proton H4 of pyridinium moieties (8.6 ppm) compared to the sum of the anomeric protons of xylose.

The formulas and the maximum, minimum and mean values of ten batches of originator SP are reported in Table 2.2.

Ratio	Formula	Min	Max	Average
3-*O*-acetylation	$$\dfrac{CH_3\ 3OAc}{OCH_3 MGA\ (xyl\ Ac)\ and\ OCH_3\ MGA\ (Xyl)}$$	0.31	0.41	0.36
3-*O*-acetylation bis	$$\dfrac{OCH_3 MGA\ (xyl\ Ac)}{OCH_3 MGA\ (xyl\ Ac)\ +\ OCH_3\ MGA\ (Xyl)}$$	0.42	0.58	0.47
Overall degree of acetylation	$$\dfrac{(CH_3\ all\ Ac)/3}{sum\ of\ anomeric\ xyl}$$	4.29 %	5.99 %	4.96 %
MGA amount	$$\dfrac{sum\ of\ anomeric\ MGA}{sum\ of\ anomeric\ xyl}$$	5.39 %	8.26 %	6.30 %
MGA amount bis	$$\dfrac{(OCH_3 MGA\ (xyl\ Ac)\ +\ OCH_3\ MGA\ (Xyl))/3}{sum\ of\ anomeric\ xyl}$$	7.19 %	8.76 %	7.99 %

Average degree of polymerization	$\dfrac{\text{sum of anomeric } xyl}{\text{sum of anomeric RE xyl}}$	8.56	7.56	7.90
Pyridinium	$\dfrac{H4\,Py}{\text{sum of anomeric } xyl}$	2.07 %	2.96 %	2.55 %

Table 2.2. ¹H-NMR integral ratios. Sum of anomeric xyl = xyl RE α *100/35 + xyl int/NR +(CH₃ 3-OAc)/3
RE = reducing end, NR = not reducing end.

These data are not absolute determinations, but relative quantifications which can be useful to quantitatively compare the ¹H-NMR spectra of commercial sulfated polysaccharide and the sample prepared during the activity of research and development.

2.2c Study of reproducibility of commercial SP by NMR

As mentioned before, we extracted few batches of capsules of Originator SP and compared the proton spectra (in Figure 2.11 the overlaid spectra of 6 batches of SP).
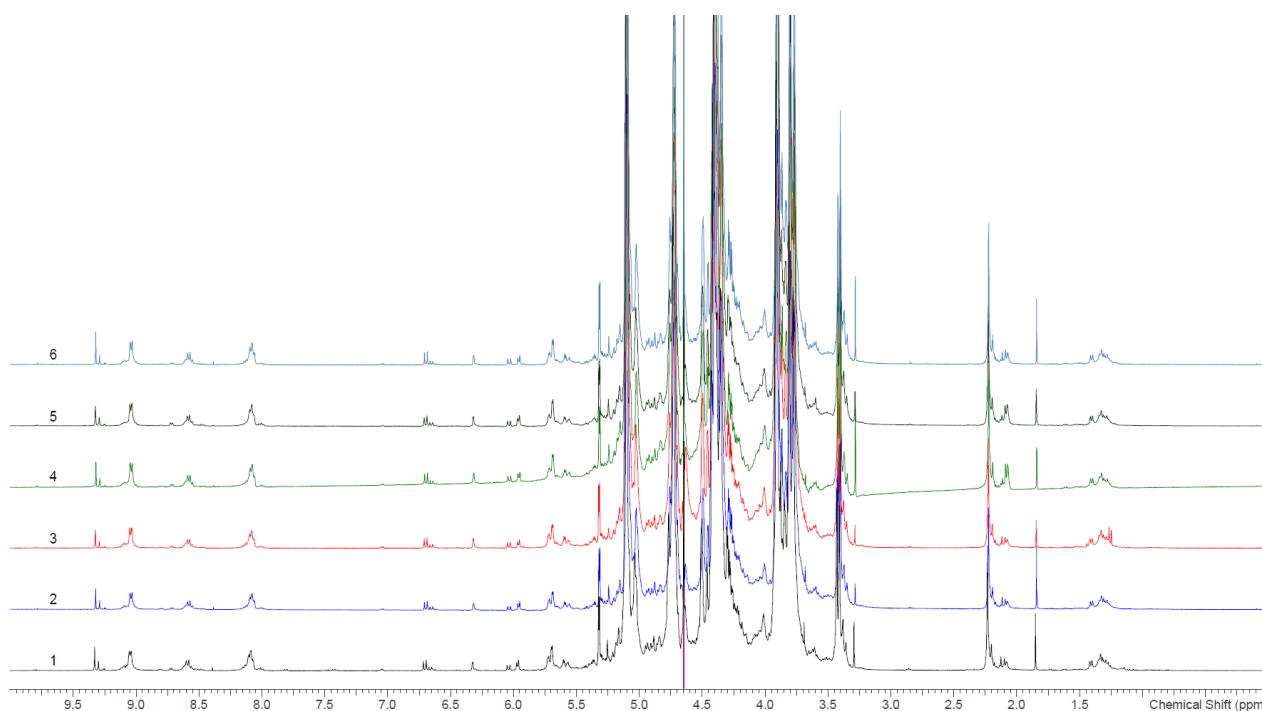
Figure 2.11. Overlay of ¹H-NMR spectra of 6 batches of commercial sulfated polysaccharide.

Looking at the proton spectra the different batches are all very similar. To compare in more detail and to better visualize the difference among the lots of SP we used Principal Component Analysis (PCA). PCA is one of the most important chemometric method useful to represent the variation of a matrix of data.[6] Indeed, when you have a huge amount of data it is complicated to analyze all the information and to catch differences among the batches. PCA can provide a solution to this

problem by building a new set of coordinates (principal components) which maximize the information contained in the matrix of data. The new coordinates are linear combination of the original variables and the coefficients of correlation (loadings) allow to move from the original variable to the principal components and vice versa. It is also possible to calculate the variance associated to each principal component in order to understand how much variation and information is described by every new coordinate. It is clear that a good set of new coordinates (principal components) has a very high percentage of associated variance.

In our case, the ^{1}H-NMR spectrum of SP is an object which can be represented as a point in a multidimensional space having one axis for each of the n-variables that define the object. Therefore, we performed PCA through Matlab to reduce the number of dimensions and to visually observe the variation among different batches of originator SP. In Figure 2.12 on the left there is the graphical representation obtained with seven batches. It is possible to define a space, highlighted in green, which contains all the commercial lots. The aim will be to prepare a sample so similar that it can fit in the green space. In Figure 2.12 the variance associated to each principal component is shown and the new space is able to describe about 84% of the variation.
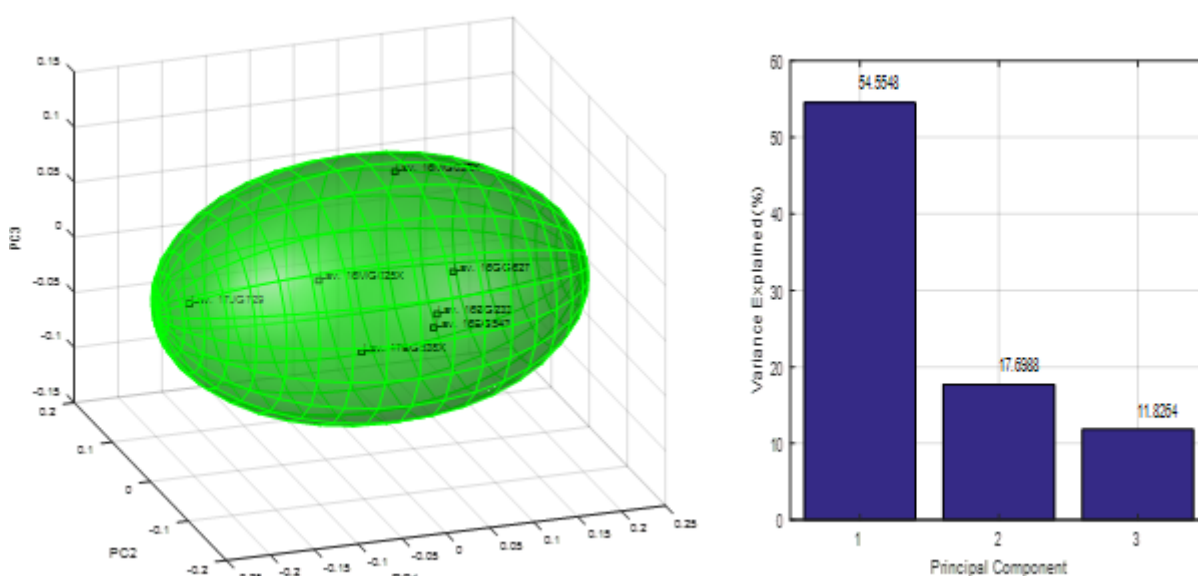
Figure 2.12: Principal component analysis (PCA) of seven batches of commercial sulfated polysaccharide.

2.2d Monosaccharides composition by NMR

The sulfated polysaccharide is mainly constituted of xylose units and ramifications of 4-*O*-methylglucuronic acid, but other monosaccharides are present as minor components. To study monosaccharide composition the product was hydrolyzed to monosaccharide units and the HSQC spectrum of hydrolyzed product was compared to spectra of pure monosaccharides. The

hydrolysis was carried out with two different methods. The first one was performed on the sulfated product using hydrochloric acid in water at 100°C. The second hydrolysis was carried out in two steps: 1) *O*-desulfation of SP by acidification with Amberlite IR120H$^+$ and heating in DMSO/MeOH at 80°C; 2) hydrolysis of the desulfated product with trifluoroacetic acid at 100°C. The results obtained with the two methods were very similar.

In Figure 2.13 the HSQC anomeric region (in red hydrolysis performed on desulfated sample and in blue hydrolysis run on sulfated product) is shown. We can notice that the hydrolysis was not complete and probably a dimeric structure of xylose bound to methylglucuronic acid xyl(MGA) is present. Indeed, the signal at 5.3-100.7 ppm is typical of the anomeric proton of MGA bound to a xylose unit of a polymeric (e.g. glucuronoxylan) or oligomeric chain. The presence of these species was then confirmed by direct infusion mass spectrometry. The disaccharides are highly stable and a further step of hydrolysis with trifluoroacetic acid was not able to cleave the dimers into the monosaccharide units.
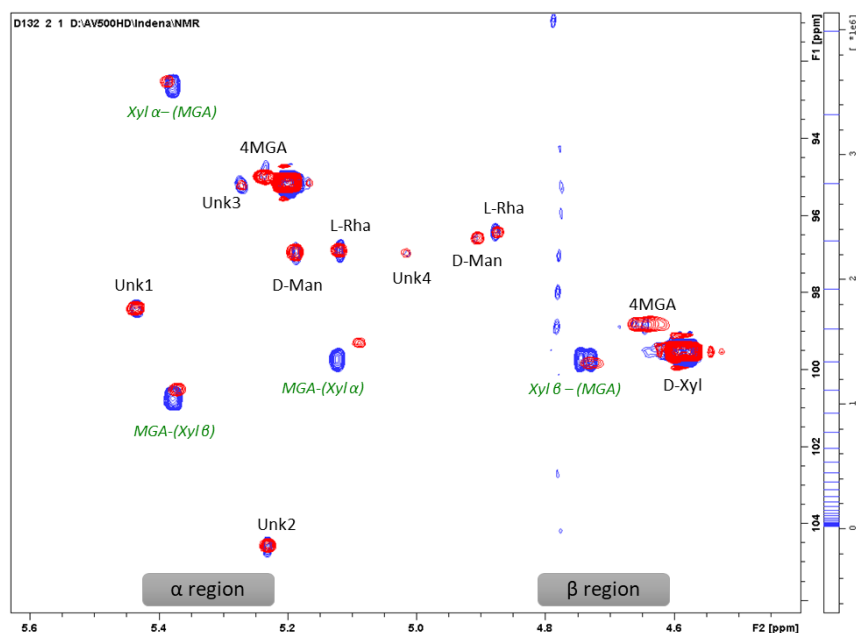
Figure 2.13. HSQC anomeric region of hydrolyzed samples. In red the sample from hydrolysis on desulfated product and in blue the sample from hydrolysis on sulfated product.

The signals of hydrolyzed samples were compared with pure samples of:

- D-Xylose (^{1}H/^{13}C α anomer = 5.20/95.1 ppm; ^{1}H/^{13}C β anomer = 4.59/99.5 ppm)
- MGA (^{1}H/^{13}C α anomer = 5.24/95.0 ppm; ^{1}H/^{13}C β anomer = 4.65/98.8 ppm)
- L-Rhamnose (^{1}H/^{13}C α anomer = 5.12/96.9 ppm; ^{1}H/^{13}C β anomer = 4.87/96.5 ppm)
- D-Mannose (^{1}H/^{13}C α anomer = 5.18/97.0 ppm; ^{1}H/^{13}C β anomer = 4.9/96.7 ppm)
- L-Arabinose (^{1}H/^{13}C α anomer = 5.24/95.5 ppm; ^{1}H/^{13}C β anomer = 4.51/99.8 ppm)

124

- D-Glucose (^{1}H/^{13}C α anomer = 5.24/95.1 ppm; ^{1}H/^{13}C β anomer = 4.65/98.9 ppm)

- D-Galactose (^{1}H/^{13}C α anomer = 5.27/95.2 ppm; ^{1}H/^{13}C β anomer = 4.59/99.4 ppm)

The comparison allowed to identify D-xylose, 4-*O*-methylglucuronic acid, L-Rhamnose and D-Mannose residues. Anomeric protons of MGA have chemical shifts similar to those of D-Glucose so that they may be easily misinterpreted. The absence of glucose in the hydrolyzed sample was confirmed by the lack of Glc H/C 6α signal at 3.85/63 ppm in the spectrum. HSQC integrals allow to relatively quantify the abundance of each sugar residue. It was found that xylose is about 90% and MGA 5% of the total sugars. Rhamnose (about 1.2%) and Mannose (about 0.9%) are minor components.

2.3 Determination of Sulfate content

API target is a highly sulfated polymer and hence it is important to determine the amount of sulfate groups present in the API. Three different analytical methods were used. The first one is based on conductimetric titration. In 1975 Casu and Gennaro described a conductimetric method to determine sulfates and uronic acids in heparins and other mucopolysaccharides.[7] According to this method, a sulfated polysaccharide is transformed into its acid form using a cation exchange resin (amberlite IR-120(H$^+$)) and then titrated by addition of sodium hydroxide solution. Titration of the acid form with strong base yields clearer end-points and more accurate results than titrating the salt with a strong acid. However, it is taken from granted that running through the resin does not alter the sample. With this method the titration curve has the shape represented in Figure 2.14. At the beginning the conductivity is high because of the contribution of mobile protons in solution. Then, conductivity decreases until a first neutralization point is reached. This corresponds to the neutralization of the sulfonic acid protons. The conductivity remains the same during the further neutralization of uronic acids. Once the latter are neutralized, the conductivity increases due to the excess of OH$^-$ ions.
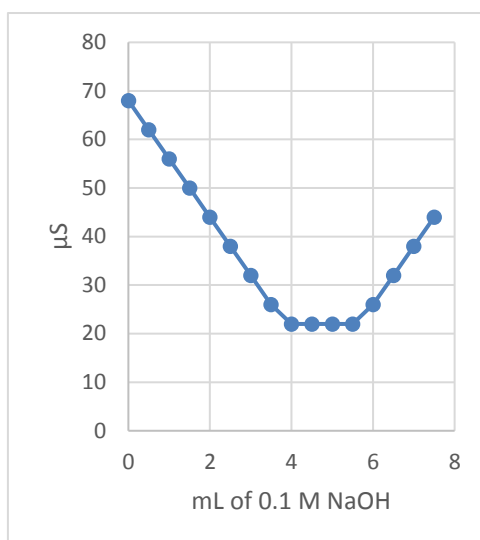
Figure 2.14. Typical conductimetric titration curve of a sulfated polysaccharide

Therefore, by a simple titration it is possible to determine the amount of sulfate groups and uronic acid units in the sulfated polysaccharide (Table 2.3). The amount of sulfate was expressed as sulfur and sulfate w/w percentage and is respectively 16.6% and 49.8%.

Sample Weight (mg)	1st end-point (mL 0.1M NaOH)	S % (w/w)	SO_3^{2-} % (w/w)	2nd end-point (mL 0.1M NaOH)	MGA % (w/w)
60.4	3.13	16.6	49.8	3.20	2.5

Table 2.3 Sulfate and MGA content of sulfated polysaccharide by conductimetric titration

The obtained values can be nevertheless affected by the presence of free sulfate groups and free acetates in the sample.

An alternative method for the determination of sulfate groups (bound and free) is ICP-MS (inductively coupled-mass spectrometry). The ICP-MS system is made up by four main components: sample introduction, ICP torch, interface and mass separation compartment (Figure 2.15). The sample introduction transforms the liquid or solid sample in aerosol or gaseous form. This can be achieved by nebulization via Argon gas flowing or by electrothermal vaporization. The second component is the ICP torch where the aerosol is atomized and ionized thanks to high temperature (6000 - 7000 K) of the Plasma, an electrical conducting gas mixture. Plasma is generated when Argon is ionized by a spark and passed through a Radiofrequency (RF) generator coil, which accelerates the cations and electrons increasing the collisions with other Argon molecules and hence increasing the temperature until 6000 K. This system allows ionization in a chemical inert environment preventing possible oxidations phenomena. After the torch, the ionic stream flows into an interface region which is fundamental to reduce temperature and pressure of

the gas before the sample enter the MS portion (mass separation device and ion detection). In this instrument the MS component is usually characterized by a quadrupole region and a detector with 26 dynodes.
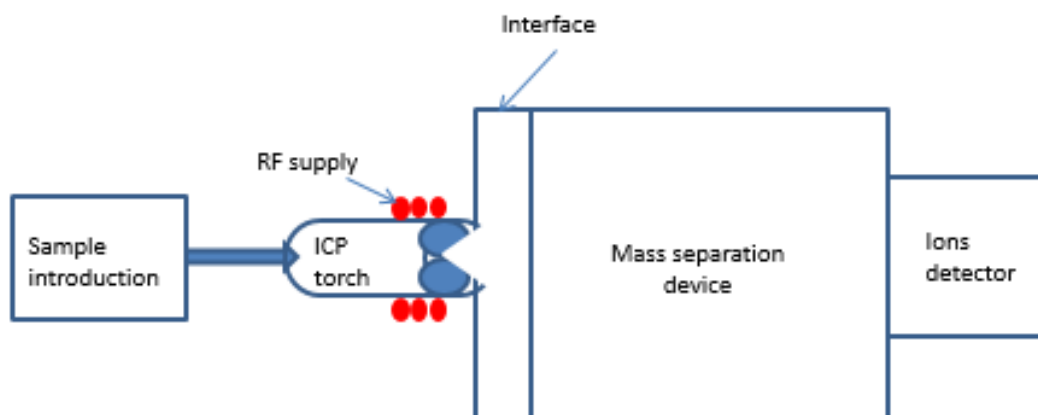
Figure 2.15. Scheme of ICP-MS ions

ICP-MS analysis allows to determine the sulfur content of the sulfated polysaccharide, which was found to be about 15.5% (Table 2.4). Like the conductimetric titration, this method is not able to discriminate between sulfur of free sulfates and sulfur of bound sulfate groups. ICP-MS also allows to determine sodium content (about 12%), which is present as counter-ion of sulfate groups, and to identify other eventual metal contaminants (e.g. barium) deriving from the process or the biomass.

A third method to quantify the sulfate groups present in the SP is by using ionic chromatography. It exploits a resin with ionic sites in the stationary phase which is able to separate ions according to their affinity. The combination of ionic chromatography and a conductivity detector allows to quantify inorganic ions.[8] This method applied to the extracted sulfated polysaccharide allows to quantify the free sulfates, which are 0.6% (corresponding to 0.2% of sulfur content). The analysis can also be carried out on the sample after complete hydrolysis and combustion to cleave the sulfates bound to sugar chains and hence to determine the overall amount of sulfates (50.4% corresponding to 16.8% of Sulfur). The difference between the total sulfates and the free sulfates yields the amount of bound sulfate groups, which is 49.8% (16.6% of sulfur). Therefore, ionic chromatography is the only method able to discriminate between the free and bound sulfate groups. On the other hand, all the three methods (conductimetric titration, ICP-MS and ionic chromatography) can determine the overall content of sulfates and the results are similar and consistent (Table 2.4).

	Titration	ICP-MS	Ionic chromatography
Sulfate % (w/w)	49.8	46.5	50.4
Sulfur % (w/w)	16.6	15.5	16.8

Table 2.4. Sulfate and Sulfur content of a batch of commercial sulfated polysaccharide.

2.4 Capillary Zone Electrophoresis (CZE)

An important analytical technique for the characterization of the sulfated polysaccharide (SP) is capillary zone electrophoresis (CZE) Electrophoresis is a process of ions separation based on their movement through a fluid under the influence of an applied electric field. Capillary zone electrophoresis is a type of electrophoresis where the separation occurs in a capillary. CZE instrument can have two possible settings according to the polarity of the electrodes in the inlet and outlet vials. In our case, because SP species are anionic, the instrument is set in "inverted polarity" mode (cathode in the inlet vial and anode in the outlet vial), as represented in Figure 2.16.
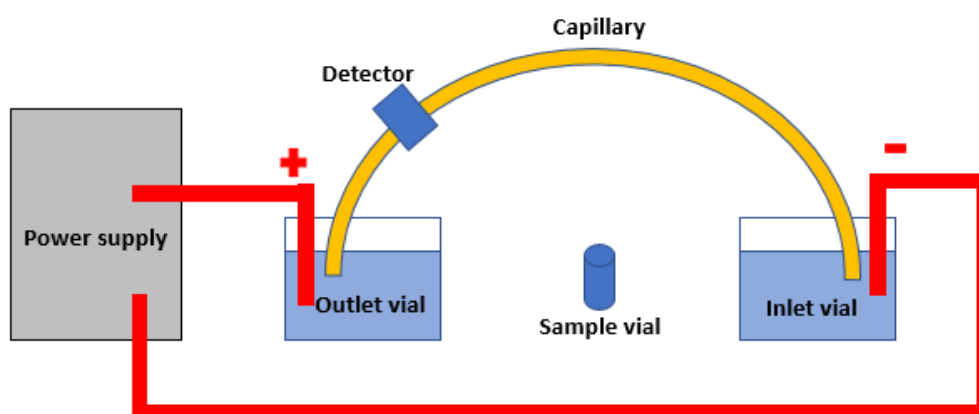
Figure 2.16. Schematic representation of CZE instrument

The capillary, which is made of a chemically inert and UV transparent material, is filled with a carrier electrolyte, also called background electrolyte (BGE) which has good buffer capacity, low background absorption and low mobility. Moreover, the capillary has an on-line detector directly placed along the route followed by the electrolytes.

A key process in electrophoresis is electroosmotic flow (EOF), which occurs because of the presence of a charged layer, the zeta potential, on the surface of the capillary. Indeed, the fused silica capillary is treated with sodium hydroxide solution ionizing silanol groups and bringing to generation of a double layer. When a voltage is applied, the negative ions migrate towards anode or positive pole. Since they are solvated, the buffer is dragged in the same direction. The

electroosmotic flow allows the migration of not only anions, but also of neutral and cationic species. EOF affects directly the measured mobility of the species in the solution, defined as apparent mobility (μ_a), which is given by the sum of the electroosmotic mobility (μ_{EOF}) and the effective mobility (μ_e).

$$\mu_a = \mu_{EOF} + \mu_e$$

The electroosmotic flow can be simply measured by injecting a neutral species in CZE and evaluating its mobility, which is only driven by electroosmotic flow. On the other hand, the charged species have a proper mobility in the solution, which is given by the balance between the electrical force and the frictional force.

$$\mu_e = \frac{q}{6\pi\eta r}$$

According to this equation, the electrical mobility (μ_e) is directly proportional to charge to size ratio (r = radius of ion) and the size of a molecule depends on the molecular weight, 3D structure and degree of solvation. Thanks to different charge to size ratio and hence different mobilities it is possible to separate the species composing the analyte. The final result of CZE analysis is an electropherogram, which is very similar to an HPLC chromatogram where the species do not have different retention time, but different migration time.

The main issue of CZE analysis of the sulfated polysaccharide is associated with the lack of chromophores. In the past several analytical methods of carbohydrates were based on derivatization of terminal semi-acetal by reductive amination or condensation with chromophores.[9,10] However, the derivatization step can cause a change in the composition of the sample and can affect the separation of the species. An alternative approach is indirect detection, which uses a BGE with a high extinction coefficient and detect the migration of analyte species through a reduction of the absorbance of the medium. In this case the choice of the buffer is crucial to allow detection and at the same time to guarantee good resolution and sensitivity. A suitable buffer can be prepared with 1,2,4-tricarboxylic acid adjusting pH to 4.9 and detecting at 217 nm.[11] A typical electropherogram of the sulfated polysaccharide obtained with this method is reported in Figure 2.17.
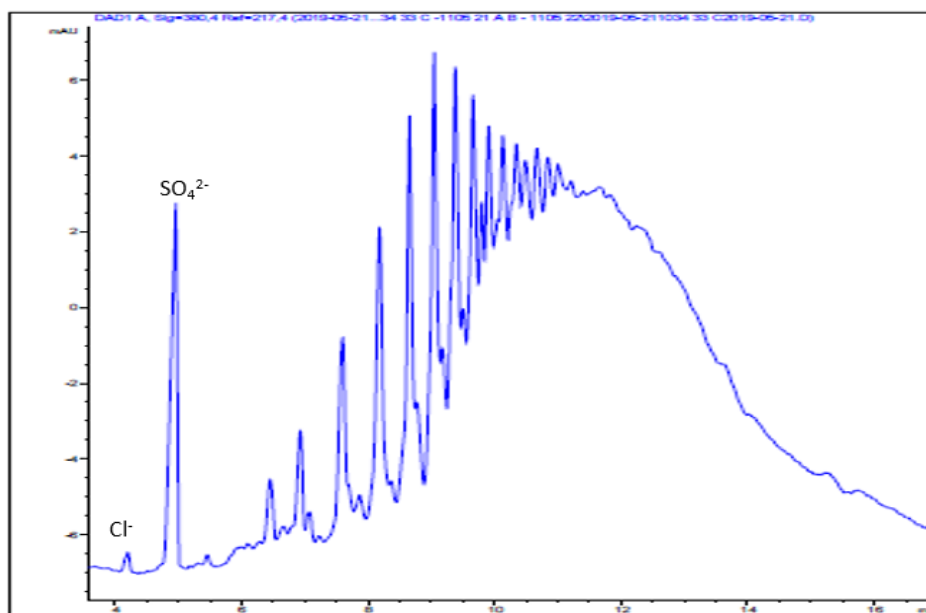
Figure 2.17. Typical electropherogram of commercial sulfated polysaccharide

As shown in Figure 2.17, at the beginning of the electropherogram there are two peaks which are related to the presence of inorganic salts (chlorides and free sulfates). The nature of this peaks was unequivocally clarified by spike experiments with NaCl and Na_2SO_4. On the other hand, it is not possible to detect a peak related to free acetate, whose presence is shown by ^{1}H-NMR spectrum, possibly because of the buffer pH. The other peaks of the electropherogram are related to the different carbohydrate chains. CZE does not give any structural information about the composition of the sulfated polysaccharide, but returns a fingerprint typical of this product. Analytical fingerprints are very important to identify and to assure the quality of complex mixture (polymers, plant extracts, natural or biotechnological products).[12] In our case, we performed CZE analysis of ten batches of Originator sulfated polysaccharide. Visually the profiles are always the same indicating consistency of the originator in the production of SP. On the other hand, the substandard SP commercialized in Latin America and India has a different pattern of peaks in the electropherogram, as shown in Figure 1.4 (paragraph 1.2). This means that CZE analysis plays an important role in the definition of the quality of the product. It is hence necessary to prepare a product with a very similar CZE profile in order to obtain an authorization for commercialization from the US authorities.

2.5 Liquid Chromatography-Mass Spectrometry (LC-MS)

Another important analytical technique for the characterization of the sulfated polysaccharide is liquid chromatography coupled to mass spectrometry. In principle this technique is similar to CZE in providing a fingerprint of the product, but it is also able to provide structural information. The main issue of LC-MS applied to charged polymers (e.g. heparins or low molecular weight heparins) is to find a chromatographic method able to separate the oligomers efficiently and that can be coupled to mass spectrometry. Literature describes a GPC-MS method used for the analysis of enoxaparin, a low molecular weight heparin, which suffers of low resolution.[13] An interesting alternative is the use of ion-pair reverse phase (IRRP) chromatography which is often used for charged molecules such as nucleic acids. Indeed, very polar molecules are not retained by stationary phases of low polarity and hence are not separated by classical reverse chromatography. The addition of an ion-pair agent in the mobile phase makes the analytes more lipophilic, increasing the interaction with the stationary phase. IRRP chromatography can be coupled to ESI-MS which provides very mild ionization conditions preventing possible fragmentations which simplify the interpretation of the spectra. IRRP-ESI-MS have been used for the analysis of highly sulfated heparin after enzymatic depolymerization[14,15] and for the characterization of Low Molecular weight heparins.[16] Because of the structural similarity of heparins with our target product, we started from these examples to develop an IRRP-ESI-MS method for the analysis of SP. The choice of the ion pair agent is crucial and should be a compromise between amines containing low alkyl chain, which enhances the volatility and hence the compatibility with MS, and long alkyl chain necessary to make the analyte more lipophilic allowing the chromatographic separation. In our case a good compromise is represented by dibutylamine (DBA).

IRRP-ESI-MS was performed on ten different batches of originator sulfated polysaccharide. The total ion current (TIC) of all the samples are shown in Figure 2.18. It is evident that ten batches have almost identical profiles and once again it is a proof of the high reproducibility of the originator process.
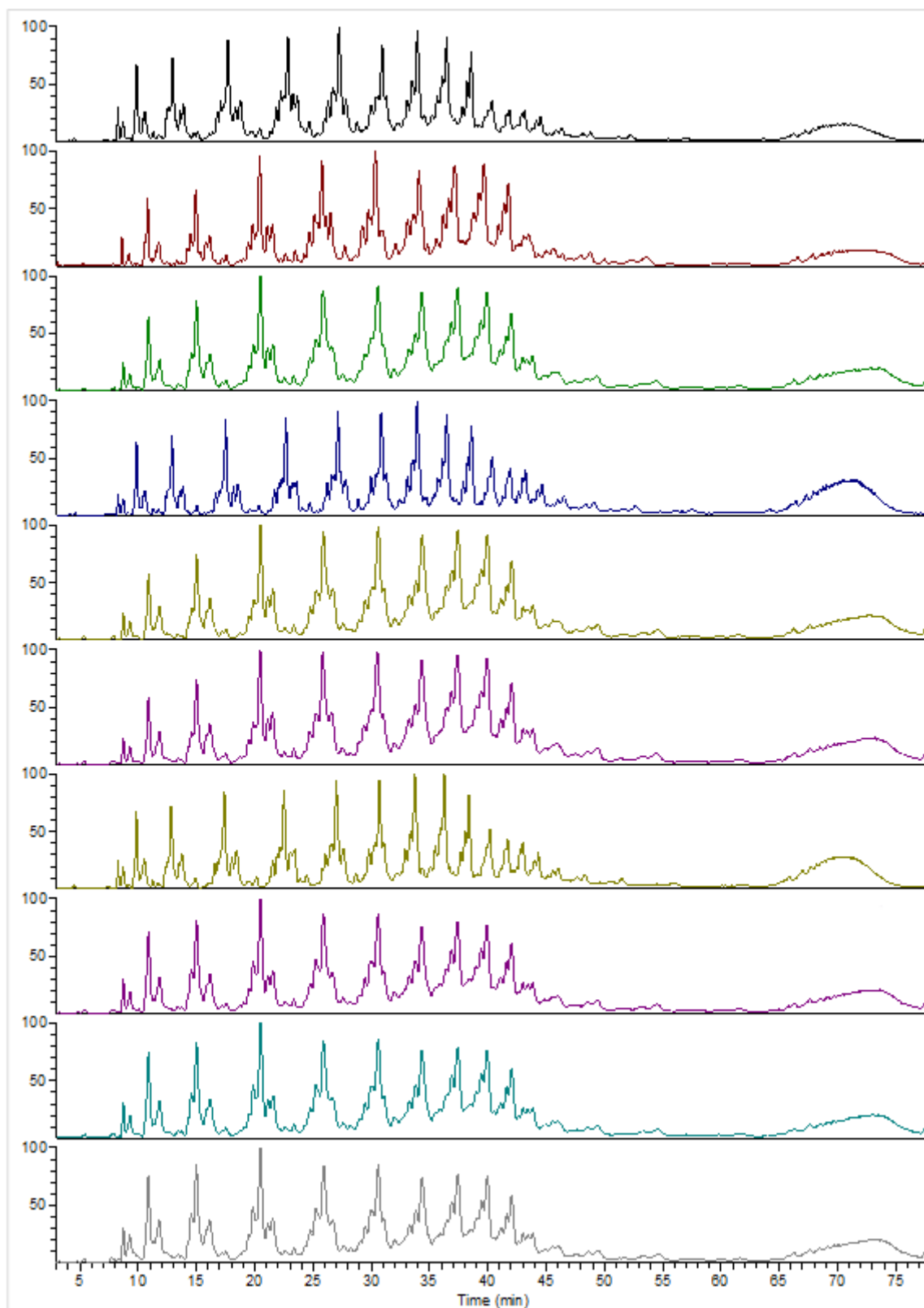
Figure 2.18. Total Ion currents (TICs) of ten batches of commercial sulfated polysaccharide

The total ion current of the HPLC chromatogram of the sulfated polysaccharide is a characteristic fingerprint of the product which will play a key role in the demonstration of sameness of a possible generic version of SP *versus* the Originator SP. The deconvolution of the signal is not

simple and our approach was to prepare a data set with m/z of the possible species and then look for them in the LC-MS chromatogram using Selected Ion Monitoring. We investigated the presence of the following species:

- Xylose chains with different degree of sulfation ($Xyl_n S_{2n+2}$; $Xyl_n S_{2n+1}$; $Xyl_n S_{2n}$; $Xyl_n S_{2n-1}$)

- Xylose chains with a ramification of 4-*O*-methylglucuronic acid and different degree of sulfation ($Xyl_n MGA S_{2n+2}$; $Xyl_n MGA S_{2n+1}$; $Xyl_n MGA S_{2n}$)

- Xylose chains with a ramification of 4-*O*-methylglucuronic acid, an acetyl group and different degree of sulfation ($Xyl_n MGA Ac S_{2n+2}$; $Xyl_n MGA Ac S_{2n+1}$; $Xyl_n MGA Ac S_{2n}$)

- Xylose chains with a pyridinium unit and different degree of sulfation ($Xyl_n S_{2n+2} Py_1$ - OH; $Xyl_n S_{2n+1} Py_1$ - OH; $Xyl_n S_{2n} Py_1$ - OH)

- Xylose chains with a dehydration and different degree of sulfation ($Xyl_n S_{2n+1}$ -H_2O; $Xyl_n S_{2n}$-H_2O; $Xyl_n S_{2n-1}$ -H_2O)

- Xylose chains with an additional methyl group (Rhamnose unit) and different degree of sulfation ($Xyl_n S_{2n+2}$ Me; $Xyl_n S_{2n+1}$ Me; $Xyl_n S_{2n}$ Me)

- Xylose chains with an hexose sugar (Gal/Glu/Man) and different degree of sulfation (Xyl_n Glu S_{2n+2} ; Xyl_n Glu S_{2n+1}; Xyl_n Glu S_{2n}).

For each species we calculated the mass from a degree of polymerization (DP) of 1 to 14 (n = 1, 2, 3,…14). Then, we calculated for every DP the MW of the oligomers with all the possible addition of dibutylamine (DBA) units as counter-ion of sulfate groups (e.g for a DP 4 the mass of the species with 1 DBA, the species with 2 DBA, the species with 3 DBA,..). Finally, for each adduct with DBA we calculated the m/z assuming the loss of 1 DBA and the generation of mono-charged ion, the loss of 2 DBA and the generation of a double-charged ion, the loss of 3 DBA and the generation of a triple-charged ion, etc. An example of this calculation for the species ($Xyl_n S_{2n+1}$) at DP = 4 is reported in Scheme 2.1 to show the principle used for the generation of the dataset.

Species	DP	Oligomer	Mass
Xyl_nS_{2n+1}	4	Xyl_4S_9	1265.8

1 DBA → $Xyl_4S_9DBA_1$ | 1349.9 → ...

2 DBA → $Xyl_4S_9DBA_2$ | 1524.1 → ...

3 DBA → $Xyl_4S_9DBA_3$ | 1653.2

4 DBA → $Xyl_4S_9DBA_4$ | 1782.4 →

z	Ion	m/z
1	$Xyl_4S_9DBA_3^{1-}$	1652.2
2	$Xyl_4S_9DBA_2^{2-}$	696.5
3	$Xyl_4S_9DBA_1^{3-}$	377.9

5 DBA → $Xyl_4S_9DBA_5$ | 1911.5

6 DBA → $Xyl_4S_9DBA_6$ | 2040.7 → ...

7 DBA → $Xyl_4S_9DBA_7$ | 2169.9 → ...

8 DBA → $Xyl_4S_9DBA_8$ | 2299.0

9 DBA → $Xyl_4S_9DBA_9$ | 2428.2 →

z	Ion	m/z
1	$Xyl_4S_9DBA_8^{1-}$	2298.0
2	$Xyl_4S_9DBA_7^{2-}$	1019.3
3	$Xyl_4S_9DBA_6^{3-}$	593.1

Scheme 2.1. Example of calculation for xylose chain with 2 sulfates per xylose and 4 xylose units. DBA= dibutylamine.

Once the dataset was completed, we started to look for the mass related to the main peaks in the calculated table and vice versa. The comparison allowed to identify the main species present in the commercial sulfated polysaccharide and then for each species we extracted its ion currents for the different degree of polymerization (Figure 2.19). The most intense peaks in TIC are related to oligomers made up by xylose units completely sulfated but one position, probably the anomeric hydroxyl group (structure **4**). When the ion current for this species (grey in Figure 2.19) is extracted, it is possible to observe two different peaks for each degree of polymerization due to the two possible anomeric configurations (α or β). A similar situation was found for the second series in green which is characterized by the presence of a pyridinium unit at the anomeric position of a xylose chain completely sulfated. The presence of these oligomers **3** was also shown by proton spectra (Figure 2.6 and 2.8, paragraph 2.2).
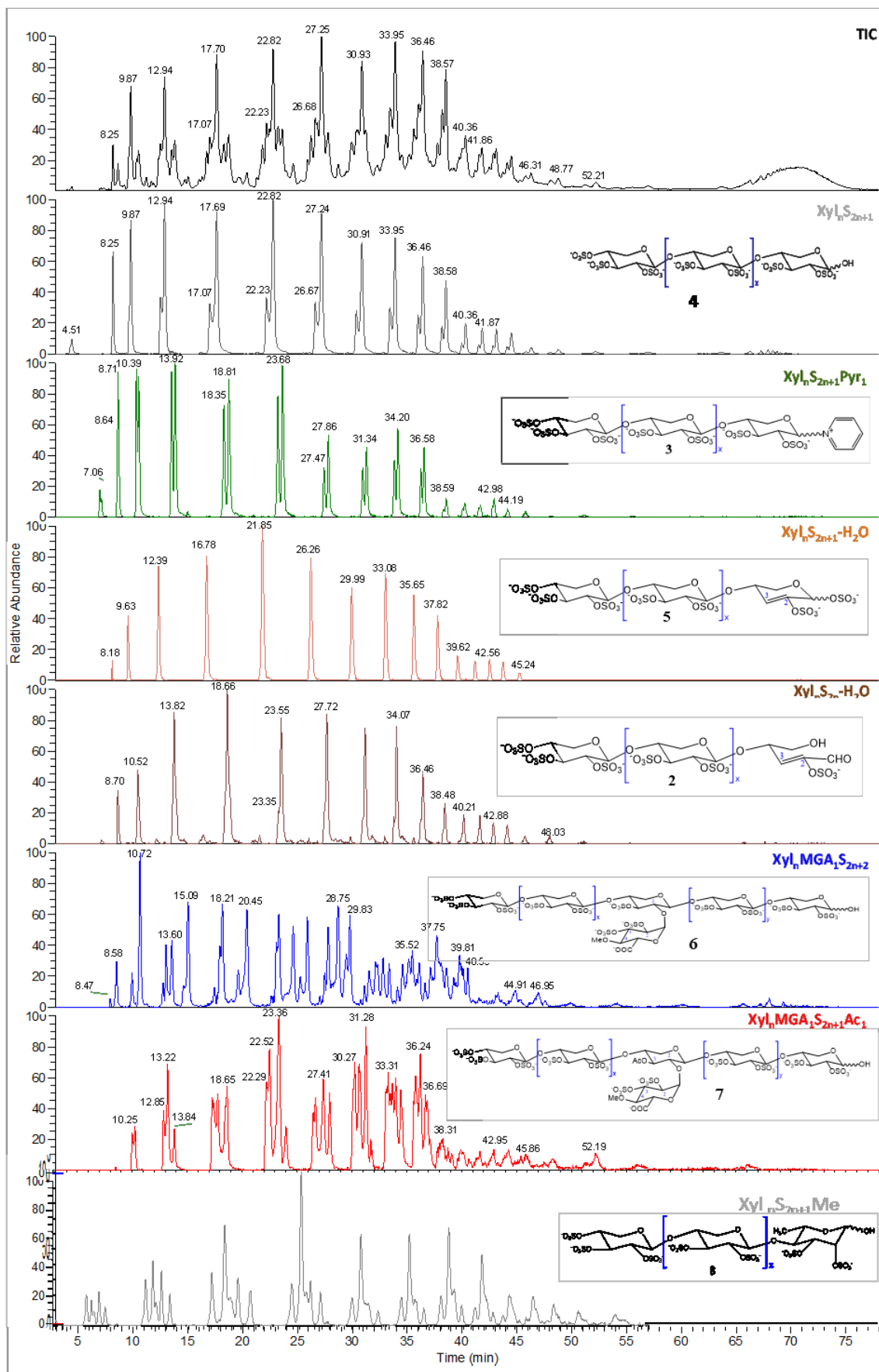
Figure 2.19. Extracted ion currents of the main species present in the commercial sulfated polysaccharide.

Two dehydrated species with different degree of sulfation were found in the commercial sulfated polysaccharide: $Xyl_n S_{2n+1}$ -H_2O and $Xyl_n S_{2n}$ -H_2O. The first one has all the possible hydroxyl groups sulfated and hence a plausible structure may be **5**, which has a double bond in a xylose unit and is formally generated by elimination of a sulfate group from species **4**. On the other hand, the second species has a hydroxyl group not sulfated and can be putatively represented as **2**, an open ring aldehyde conjugated to a double bond in position 2 and 3. The presence of this structure is supported by ^{1}H-NMR spectrum (Figure 2.6, paragraph 2.2). Further, we found two species characterized by ramification along the xylose chain with 4-*O*-methylglucuronic acid: the first one is without acetyl group (structure **6**, blue) and the second one has an acetyl group, which is located in positon 3 of xylose bearing MGA unit according to NMR analysis (structure **7**, red). The last series in grey in Figure 2.19 is a methylated species which may correspond to a chain where one sugar unit is a Rhamnose instead of xylose and all the hydroxyl groups are sulfated but the anomeric position.

We were not able to identify a series where a hexose sugar replaces a xylose unit although the NMR study on the hydrolyzed SP suggests the presence of mannose. Probably the amount of these oligomers is so low that the ion currents do not emerge from the noise.

Finally, we observed that for all the species it is possible to detect the well-defined peaks until degree of polymerization of 14. The higher molecular weight oligomers are not separated by the chromatographic system and are merged in the broad peak evident in TIC chromatogram at about 70 minutes.

2.6 Additional analytical tests

We have also evaluated the vibrational properties of the commercial sulfated polysaccharide by Fourier Transform Infrared Spectroscopy and Raman spectroscopy. The first one was performed both on the Originator SP and the substandard Indian product (Figure 2.20). With this technique the two samples look very similar. Indeed, both the spectra share exactly the same bands (e.g. at about 1200 cm^{-1} the vibration S=O band and at 980 cm^{-1} C-O-S stretching). However, we know that using other analytical techniques the two products are different; for example, NMR shows the absence of acetyl groups in the indian product and CZE analysis shows different fingerprints. Therefore, we can conclude that FTIR is not able to detect relevant differences between products.
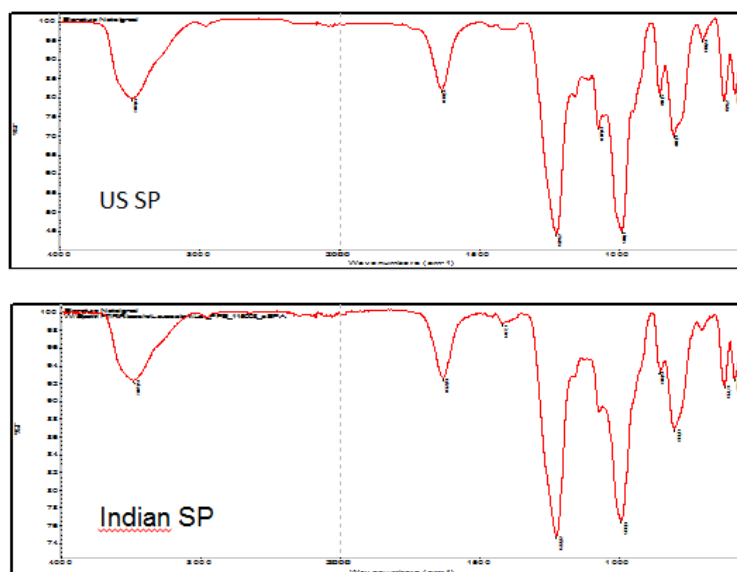
Figure 2.20. FT-IR spectra of sulfated polysaccharide (SP) from United States and India.

The same considerations are also valid for Raman spectroscopy obtained by excitation at 532 nm (Figure 2.21). Originator SP and substandard Indian product have the same characteristic bands (590 cm^{-1}, 870 cm^{-1}, 1030 cm^{-1}, 1070 cm^{-1}, 3000 cm^{-1}) indicating also in this case a low sensitivity of this technique to structural differences. Moreover, the presence of traces of fluorescent components can seriously interfere in the acquisition of the spectra and thus Raman spectroscopy is suitable only to compare white samples (slightly colored SP contain fluorescent components).[17]
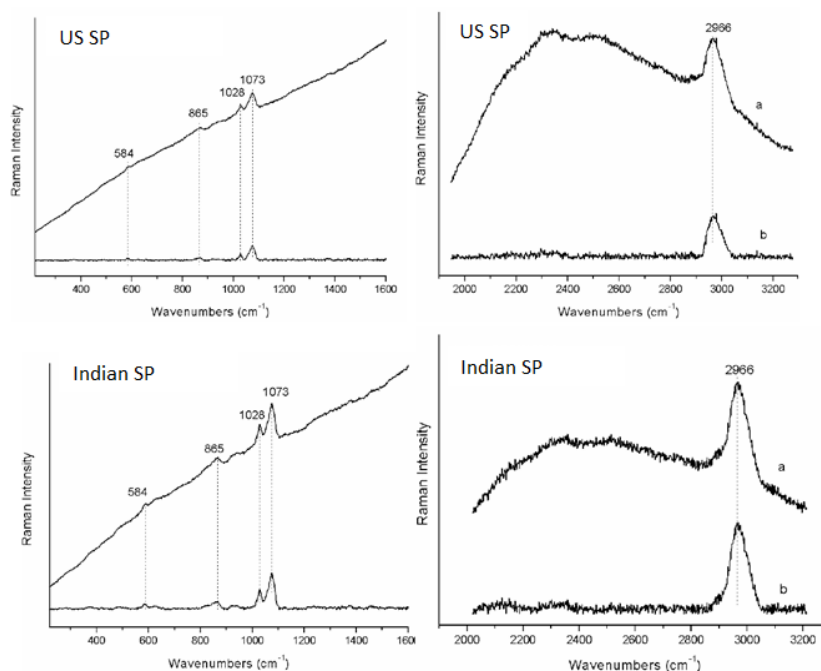
Figure 2.21. Raman spectra of the sulfated polysaccharide (SP) from United States (top) and India (bottom). On the left there is the region between 220-1600 cm^{-1} and on the right between 1900 and 3300 cm^{-1}. For each region there is the raw spectrum and the spectrum corrected for the baseline.

Finally, we investigated the optical activity of the sulfated polysaccharide with a polarimeter. This property is generally expressed as $[\alpha]_D$, the specific rotation obtained using the D-line of the Sodium lamp. The originator SP has $[\alpha]_D$ equal to -60° (10% w/v solution in water). This value does not contain any structural information, but it is another parameter useful to compare the originator's product and the samples prepared during the research activity.

2.7 Fractionation of SP

Inspired by a work published in literature,[18] we carried out preparative size exclusion chromatography (SEC) and then analyzed the different fractions in order to better characterize the Originator SP. Each fraction was analyzed by GPC-RI to determine both its averaged molecular weight and the gravimetric contribution to the product. Moreover, the fractions were analyzed by CZE and LC-MS to understand the contribute of a group of oligomers with a certain degree of polymerization to the overall electropherogram and chromatogram of the product.

The separation was achieved using Toyoperal HW 40-S, a methacrylic resin with a size exclusion limit of 3×10^3 Dalton (globular protein) and 30 µm as average particle size. The Originator SP was dissolved in water and charged on the resin (3.6 L of resin for gram of SP). The resin was eluted with water and the presence of product in the eluate was monitored by TLC. Actually, it is not possible to perform a real TLC, because the product does not run on the silica gel plate. Nevertheless, it is possible to exploit TLC to simply detect the presence of the product at the spotted site when stained with Cerium Ammonium Molybdate. We collected and freeze dried five different fractions containing SP. The amount of product in each fraction and the results obtained with GPC-RI are reported in Table 2.5.

Fraction	Amount (mg)	Mn	Mw	Mp	PD
Commercial SP	500	3348	7144	4484	2,13
A	143	6350	7584	6860	1,19
B	111	6137	7509	6068	1,22
C	82	5016	6108	4542	1,22
D	67	3654	4345	3458	1,19
E	38	2560	3343	2271	1,31

Table 2.5. Amount and GPC-RI results of the fractions obtained after preparative size-exclusion chromatography of commercial SP.

The total amounts of the five fractions is 441 mg, which stands for a column recovery of 88%. The discrepancy between what has been charged and recovered from the column, can be ascribed to

physiological and operative losses. On the other hand, it is also possible that we have not collected the extremes of MW distribution (very high and very low molecular weight oligomers).

We can observe that all the MW parameters (Mn, Mw, Mp) decrease from fraction A to fraction E and the polydispersity of the fractions is low. This is an indication of a satisfying separation where, as expected, the high molecular weight oligomers elute first and the low MW ones last. Furthermore, the chromatographic separation allows to weight the different molecular weight species. For example, all the oligomers with molecular weight above 6000 Da (fractions A and B) are about 50% of the product, while about 20% of SP species have MW below 4000 Da (fractions D and E). Finally, the inorganic salts (acetates, chlorides and sulfates), which have very low MW, eluted partially in fraction E and partially after fraction E.

As anticipated before, we also run capillary zone electrophoresis (CZE) and liquid chromatography-mass spectrometry (LC-MS) of the fractions. The profiles are reported in Figure 2.22 (on the left LC-MS and on the right CZE). LC-MS profiles confirm MW results obtained by GPC-RI because in the first fraction we observe the broader peak at 70 minutes and the peaks at high retention times, which correspond to high degrees of polymerization. On the other hand, in the last fraction the most intense peaks are at low RT and we observe that these peaks (monomer and dimers) are only present in fractions D and E, which are about 20% of the product w/w. Nonetheless, they are pretty intense in the overall LC-MS profile of SP because the low molecular weight oligomers respond better and give a substantial contribution in the LC-MS fingerprint.
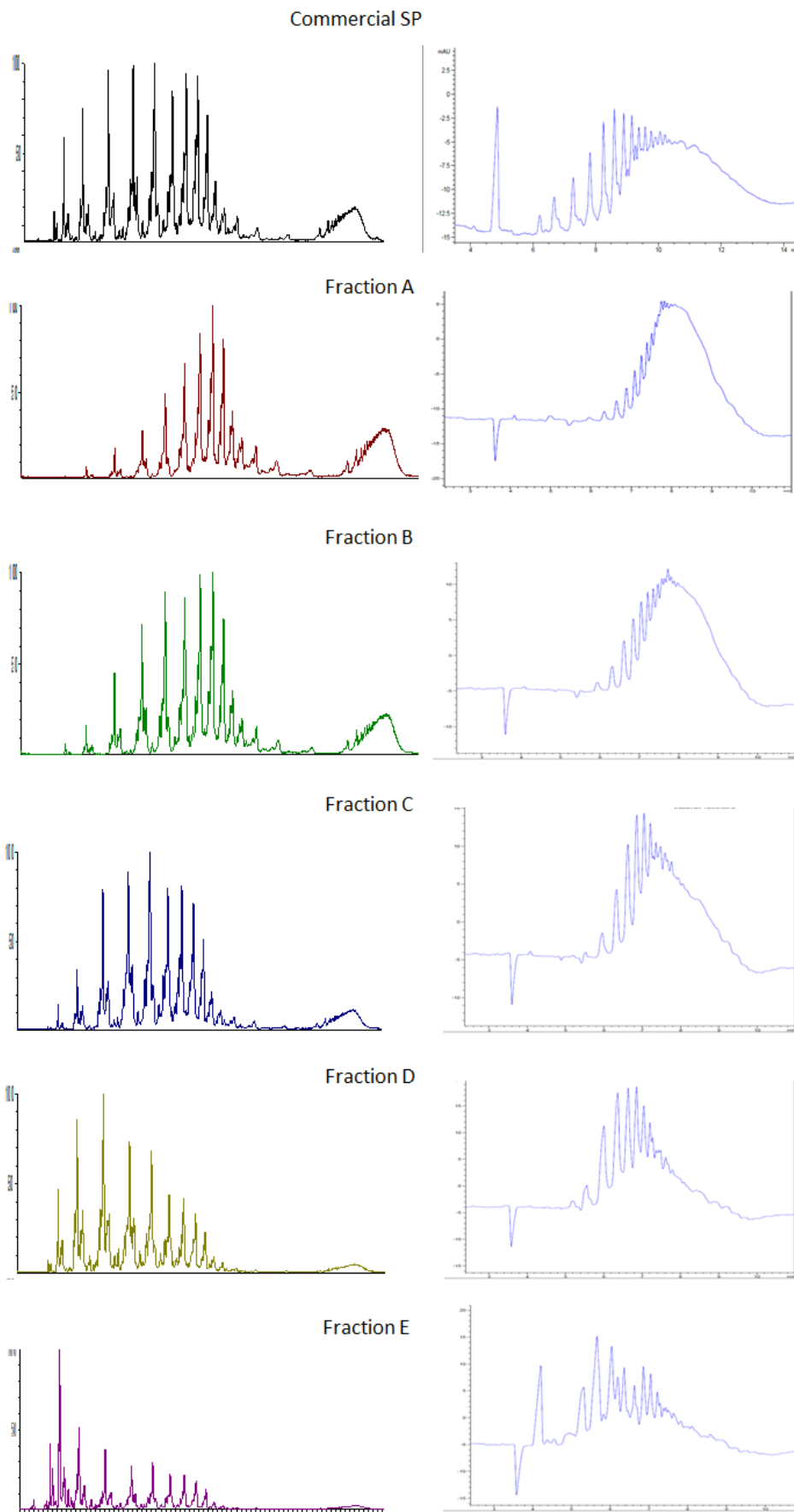
Figure 2.22. LC-MS and CZE profiles of the fractions of commercial SP

CZE analysis shows a shifting of the apex of the curve towards lower migration times as the fractions of SEC column moves to lower molecular weight. This behavior could not be given from granted before the fractionation experiment since CZE depends both from charge and volume of the species. It can be concluded that during CZE analysis the species with low MW (less charge and lower hydrodynamic volume) migrate faster than species with high MW (more charge and higher volume). Moreover, it is possible to observe that the sharp peaks characterizing the first part of the electropherogram have a higher intensity in the last fractions. This means that in CZE the low molecular weight oligomers have a key role in the definition of the fingerprint.

Finally, we can observe that the first intense peak, related to the presence free sulfates ions is present only in the last fraction. This evidence is a corroboration that the size-exclusion separation was effective because species with very low molecular weight, like an inorganic salt, were more retained.

References

[1] An Introduction to Gel Permeation Chromatography and Size Exclusion Chromatography, Agilent Technologies.

[2] A guide to multi-detector gel permeation chromatography, Agilent Technologies

[3] S. Bertini, A. Bisio, G. Torri, D. Bensi, M. Terbojevich, *Biomacromol.*, **2005**, 6, 168-173.

[4] L. De Ferra, A. Naggi, M. Zenoni, **2014,** patent appl. WO 2014/114723.

[5] L. Ahrgreen, A. de Belder, T. Malson, *Carbohydr. Polym.*, **1991**, 211-214.

[6] R. Bro, A. Smilde, *Anal. Methods*, **2014**, 6, 2812-2831.

[7] B. Casu, U. Gennaro, *Carbohydr. Res.*, **1975**, 39, 168-176.

[8] H. Small, *J. Chromatogr. A*, **1991**, 546, 3-15.

[9] C. Chiesa, C. Horvath, *J. Chromatogr. A*, **1993**, 645, 337-352.

[10] S. Honda, S. Suzuki, A. Nose, K. Yamamoto, K. Kakehi, *Carbohydr. Res.*, **1991**, 215, 193-198.

[11] M. Degenhardt, H. Benend, H. Watzig, *J. Chromatogr. Ay*, **1998**, 817,297-306.

[12] B. Schirm, H. Benend, H. Watzig, *Electrophoresis*, **2001**, 22, 1150-1162.

[13] Q. Zhang, X. Chen, Z. Zhu, X. Zhan,Y. Wu, L. Song, J. Kang, *Anal. Chem.*, **2013**, 85, 1819-1827.

[14] C. Thanawiroon, K. Rice, T. Toida, R. Linhardt, *J. Biol. Chem.*, **2004**, 279, 2608-2615.

[15] J. Henriksen, P. Roeptstorff, L. Ringborg, *Carbohydr. Res.*, **2006**, 341, 382-387.

[16] D. Li, L. Chi, L. Jin, X. Xu, X. Du, S. Ji, L. Chi, *Carbohydr. Polym.*, **2014**, 99, 339-344.

[17] F. Cabassi, B. casu, A. Perlin, *Carbohydr. Res.*, **1978**, 63,1-11.

[18] M. Degenhardt, P. Ghosh, H. Watzig, *Arch. Pharm.*, **2001**, 334, 27-29.

3. Process definition

In Chapter 2 we discussed the use of orthogonal analytical methods to characterize and identify the key features of the Originator SP. Simultaneously, we focused on the definition of a process for the preparation of a product which can be considered a generic version of the SP.

The first part of the process concerns the extraction of a partially acetylated glucuronoxylan from beech wood sawdust. For this we evaluated three different methods: chlorine dioxide delignification followed by ammonia extraction, peracetic acid delignification followed by ammonia extraction and autohydrolysis. The second part of the process involves sulfation of the polysaccharide and the purification steps (e.g. bleaching and removal of inorganic salts).

3.1 Beech wood and its composition

The starting material of the whole process is beech (*Fagus*) wood, which belongs to *Angiospermae*, known as hardwoods. The first monograph on *Fagus* was published by Shen in 1992 and divided the different species in two main subgenera according to morphology[1]. The first one is subgenus *Engleriana*, which includes three different species coming from China, South Korea and Japan (*Fagus engleriana*, *Fagus japonica* and *Fagus okamotoi*). In the second one, subgenus *Fagus*, there are ten different species and among them *Fagus sylvatica* (from Europe and Southwestern Asia) and *Fagus grandifolia* (from Eastern North America and Mexico). However, this classification was not supported by molecular phylogenetic studies performed by Manos and Stanford.[2] For the preparation of the SP, we have used *Fagus sylvatica*, which is a large tree able to reach heights up to 50 m and trunk diameter of 3 m. It has a typical lifespan of 150-200 years and it's usually harvested at the end of the vegetative session at 80-120 years of age. It is mainly located in western and Central Europe, especially in Germany (Figure 3.1).[3]

Figure 3.1. Distribution map of Fagus sylvatica

Beech wood (*Fagus sylvatica*) is made up by three main components: lignin (22%), cellulose (45%) and xylan (28%).[4] The most abundant compound is cellulose **9**, a homopolymer of β-(1-4)-linked D-glucopyranose (Figure 3.2). The hydroxyl groups in positions 2,3 and 6 are free, as demonstrated by Brawn which obtained 2,3,6-trimethyl glucose as the only product of methylation and hydrolysis of cellulose.[5] It has weight average degree of polymerization of 10000 and it is present in the cell walls organized in elementary fibrils. Each fibril contains about 40 cellulose chains and has a diameter of 35 Å.[6] The elementary fibrils aggregate originating large microfibrils with different sizes. Cellulose has several industrial applications: paper, nitrocellulose, cellulose acetate, etc.

9
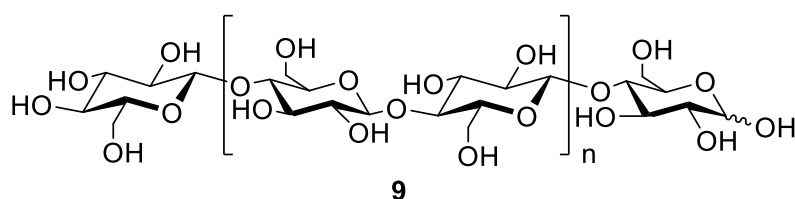
Figure 3.2. Structure of cellulose.

The cellulose fibers are surrounded by another wall polymer, lignin, which plays a key role in the mechanical strength of the plant cells and in the resistance to pathogens.[7] Moreover, lignin controls also solute transport and water content of the plant, thanks to its hydrophobicity. The structure of lignin is still not completely clear because such a complex molecule is not easy to characterize. However, it is well established that lignin is a polymer of phenylpropane residues and a putative structure **10** is represented in Figure 3.3.

144
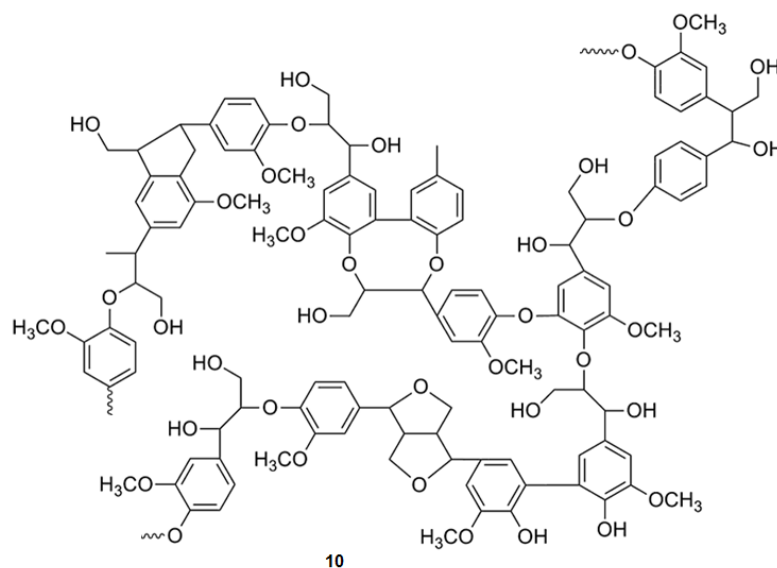
Figure 3.3. Structure of a segment of lignin.

The precursors of lignin are three derivatives of cinnamyl alcohol: coniferyl alcohol **11**, sinapyl alcohol **12** and *p*-coumaryl alcohol **13** (Figure 3.4)[8]. They are present as glucosides and are released by a β-glucosidase before their polymerization. The polymerization probably proceeds via radicals joining the monomeric precursors to the growing chain. At the end the structure is characterized mainly (about 60%) by arylglicerol-β-aryl ether bonds, but we can also find benzyl aryl ethers, phenylcoumarans, biphenyls and diaryl ethers.

11   **12**   **13**
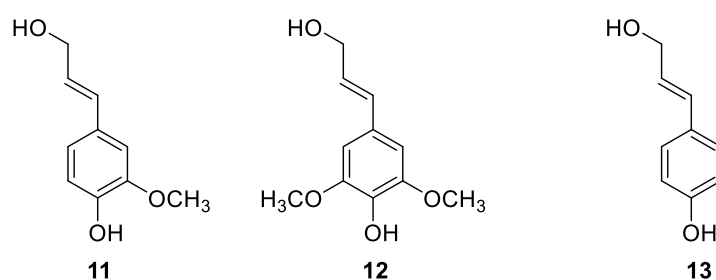
Figure 3.4. Structure of lignin precursors: coniferyl alcohol (11), sinapyl alcohol (12) and *p*-coumaryl alcohol (13).

There has been a long debate about the interaction of lignin and polysaccharides and now it is clear that the two species are both associated by physical forces and by covalent bonds.[8] Indeed, there are indications of ester, ether and glycosidic bonds between lignin and polysaccharides. For example, the oxygen of benzyl alcohol of lignin can be connected to the carboxyl group of 4-*O*-methylglucuronic acid of a xylan chain or the α position of lignin can be bound to the hydroxyl group of sugar units (xylose or glucose). The presence of lignin in beech wood and its association

to the polysaccharide matrix plays a key role in xylan extraction and, as it will be discussed in the next paragraphs, several methods of xylan extraction require prior delignification.

The third main component of beech wood is the glucuronoxylan and will be discussed in detail in the next paragraph. Finally, there are minor components (2-8%) called extractives, organic low molecular weight materials such as terpenes, phenols, carboxylic acid, mono and disaccharides and amines.[9] In the sawdust also inorganic ashes characterized by 30 different elements are present (e.g. Potassium, Magnesium, Barium).

3.2 4-*O*-methylglucuronoxylan

The component of beech wood essential for the preparation of the sulfated polysaccharide is the glucuronoxylan, which belongs to the category of hemicelluloses. Hemicelluloses are low molecular weight polysaccharides present in plants together with cellulose. In hardwoods the hemicelluloses are essentially 4-*O*-methylglucuronoxylans and glucomannans. Glucuronoxylan **14** is characterized by a main backbone of β-(1-4)-D-xylopyranosyl units decorated with 4-*O*-methylglucuronic acid (MGA) residues (Figure 3.5). MGA is attached, through its anomeric position, at position 2 of a xylose residue as described by Jones and Wise, which hydrolyzed the sawdust and among the obtained sugars found the dimeric structure of (1-2) linked (4-*O*-methyl-α-glucurono)xylose.[10]
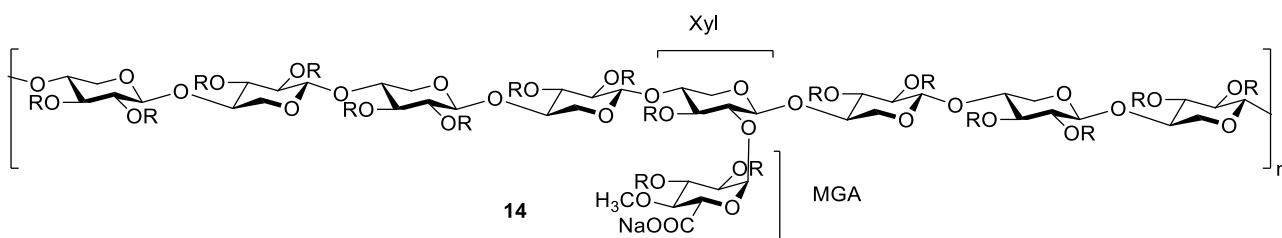
Figure 3.5. Structure of (4-*O*-methylglucurono)xylan. R = H or Ac

Several studies were performed to determine the number and the distribution of MGA ramifications along the xylose chain. Timell described the presence of a glucuronic acid each 10 xylose units[6], while Teleman reported the presence of one glucuronic acid every 15 xylose residues, based on the analysis carried out by HPAEC-PAD and NMR spectroscopy.[11] Anyway, MGA is probably not uniformly distributed, but two different domains can be identified.[12] Indeed, plant cells contain two different glucuronyltransferaeses (GUX1 and GUX2) which allow the linkage of MGA units to the xylose backbone. GUX1 favors the addition of MGA every 10 or more residues (major domain), while GUX2 decorates the backbone attacking glucuronic acid every five, six or

seven residues (minor domain). Therefore, the xylan chain is characterized by regions with spaced MGA units and highly branched regions.

Furthermore, the xylose chain features also acetyl groups located in position 2 and 3 of the xylose units with prevalence of acetylation in position 3. The amount of acetyl groups was measured by determination of acetic acid content after hydrolysis and by NMR spectroscopy. It was found that there is about one acetyl group every two xylose units.[11] According to a recent study, it is likely that there is a pattern of acetylation of alternate xylose units, especially in the major domain.[13] As discussed in Chapter 2, acetylation is a key feature of the Originator SP, which presents a prevalent acetylation at position 3 of xylose units bearing MGA in position 2. Therefore, it is essential to find a method which extracts the xylan with acetyl groups intact and preferably with acetates present mainly in the cited position.

Finally, the molecular weight of the extracted *O*-acetyl-(4-*O*-Methylglucurono)xylan was studied, finding that the results are strongly affected by the method of xylan extraction. For example, xylan extracted with alkali has a weight average molar mass of 16000[14], while xylan obtained after chlorite delignification and dimethyl sulfoxide extraction has a weight average molar mass of 11000.[11] However, in both cases the yield of extraction was below 10% and if you compare this value to the amount of xylan present in beech wood (28%) it is evident that the extraction was not complete.

3.3 Commercially available xylans and paper industry

First of all, we evaluated the sourcing of commercially available glucuronoxylans, which can be exploited for the preparation of the SP. The main advantage of using such sources could be avoiding the extraction step of the xylan from beech wood. We obtained five xylans from different producers (Senn, Megazyme, Carbosynth, Lenzing, NBS biological) and we analyzed them by ^{1}H-NMR (Figure 3.6). The proton spectra revealed immediately the complete absence of acetyl groups in all the samples because in the area between 1.8 and 2.3 ppm there is only the singlet of sodium acetate. Since acetyl groups are a key feature of the Originator SP it is not possible to use commercially available glucuronoxylan to prepare our target product.
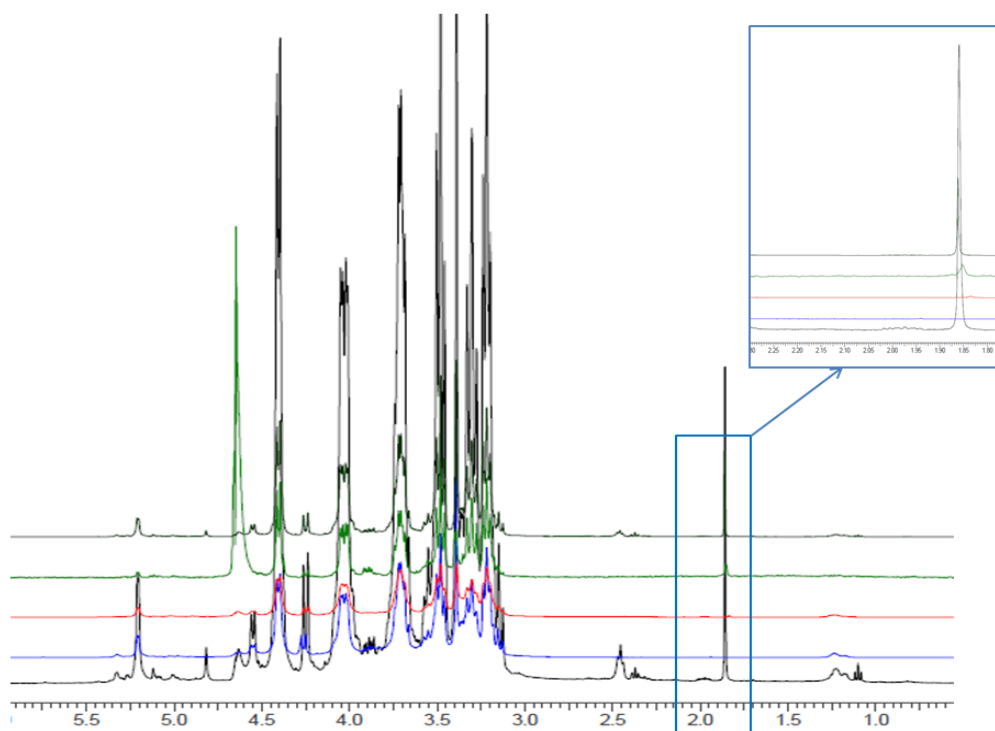
Figure 3.6. ^{1}H-NMR spectra of commercially available xylans from five different producers (Senn, Megazyme, Carbosynth, Lenzing, NBS biological). In the blue square the acetyl groups area is shown.

Another option might be exploiting the paper industry supply chain and use the pulp obtained from beech wood for the preparation of paper. The pulp is lignocellulose whose most common production process is the so-called Kraft process, which treats wood chips with water, sodium hydroxide and sodium sulfite at high temperatures (150-180°C). It is reported that it is possible to recover xylans from wood pulp using alkaline conditions, but it high NaOH and KOH concentrations (higher than 5%) are required to obtain good extraction yields.[15] The use of strong basic conditions removes completely any acetyl group present on the chain[16,6] and hence makes the xylan not suitable for the preparation of SP. Therefore, we can not use commercially available xylans and intermediates of the paper industry, but we have to extract the xylan directly from beech wood sawdust. We evaluated three different methods, which will be discussed in the next paragraphs: chlorine dioxide delignification followed by ammonia extraction; peracetic acid delignification followed by ammonia extraction and autohydrolysis. Goal of this part of the project is to obtain a glucuronoxylan mainly acetylated in position 3 of xylose units that also bear MGA moieties. The target molecular weight of the xylan should be about 1500-2500 Da, because theoretically these values allow to obtain the correct MW of final SP.

3.4 Xylan extraction: Chlorine dioxide method

The first method for the extraction of glucuronoxylan involves the delignification of beech wood with chlorine dioxide and the extraction of the glucuronoxylan from delignified sawdust.

3.4a Beech wood delignification with chlorine dioxide

The delignification step is necessary to remove lignin from beech wood in order to allow the extraction of the xylan efficiently and with mild conditions. An efficient method of delignification is the use of chlorine dioxide, a strong oxidant. Because of its complex and heterogeneous structure, the reactions of lignin with ClO_2 are not completely understood. However, it was possible to identify a main reaction pathway with the phenolic structures of lignin (Scheme 3.1).[17] The oxidant is able to abstract a H· from a phenolic hydroxyl group generating a phenoxy radical which is stabilized by resonance. Due to the delocalization of the radical, the attack of a second equivalent of chlorine dioxide can occur in *ortho* or *para* position. When ClO_2 is bound in *ortho* position, the intermediate **17** evolves into the muconic acid derivative **18**. On the other hand, if the *para* position is involved, you observe the formation of intermediate **19** which undergoes elimination generating quinoid derivatives **20.**
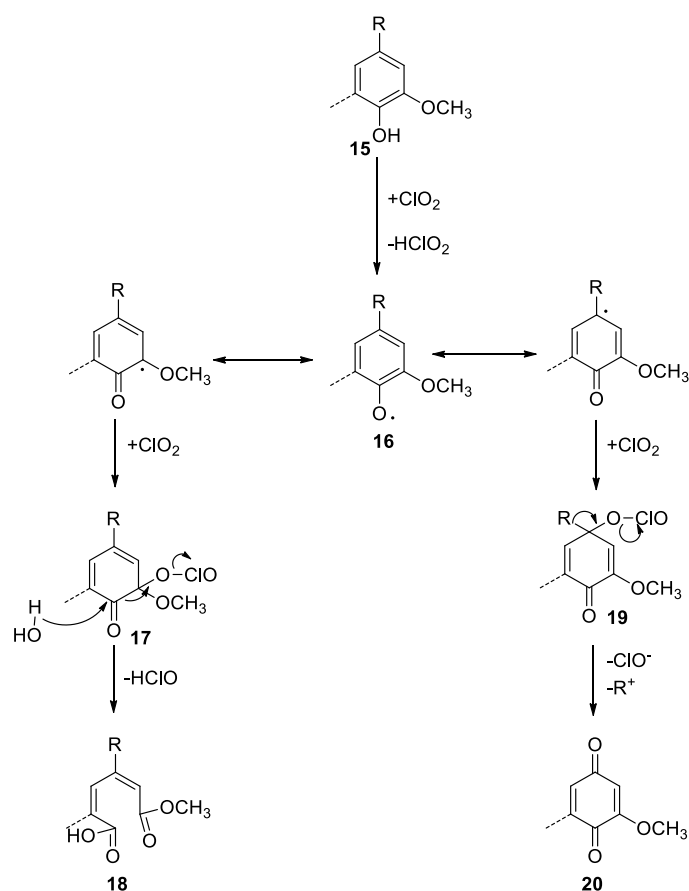
Scheme 3.1. Mechanism of oxidation of lignin phenolic structures by chlorine dioxide. R = H or CH_2OH.

Furthermore, chlorine dioxide is also able to react with lignin non-phenolic structures according to a pathway which proceeds via radical oxonium ions and provides eventually muconic acid and quinoid derivatives. The mechanism and the products are similar to those of phenolic structures, but the reaction rate is much slower.[18]

Finally, chlorine dioxide acts also on conjugated structures with a different reaction pattern (Scheme 3.2).[19] The oxidant attacks the double bond with generation of a benzylic radical in α position which evolves into epoxide **23**.
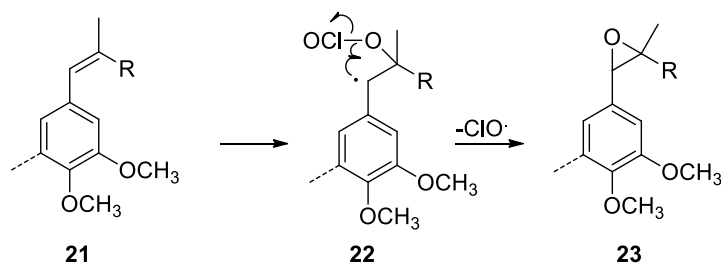
Scheme 3.2. Mechanism of oxidation of conjugated structures with chlorine dioxide.

The overall result of all these oxidation pathways is that the structure of lignin is broken into different fragments which are more soluble in water and hence can be removed from beech wood sawdust.

Chlorine dioxide is usually generated in situ from sodium chlorite and acetic acid which react to yield chlorine dioxide, sodium acetate and sodium chloride.

$$4\ AcOH + 5\ NaClO_2 \longrightarrow 4\ ClO_2 + 4\ NaOAc + NaCl + 2\ H_2O$$
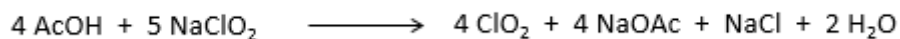
In the first lab trial we used the procedure developed by Dr. Capek (Slovak Academy of Science): the addition of AcOH and $NaClO_2$ was repeated three times (3x44 mmol of AcOH and 3x118 mmol of NaClO in 0.5 L of water in a 2L flask) at 60°C and the mixture was stirred at that temperature for an overall reaction time of 3h. The main problem of this approach is the chlorine dioxide itself, which is considered a highly explosive gas. According to Bretherick's Handbook, chlorine dioxide has a limited stability and an atmosphere with more than 10% of chlorine dioxide (0.1 bar partial pressure) is considered highly explosive.[20] Considering the reaction between acetic acid and sodium chlorite quantitative, for the experiment outlined above there is the overall formation of 132 mmol of chlorine dioxide. From the solubilities and partial pressures at 60°C data[21] we derived the relationship which correlates the solubility of ClO_2 in water and its partial pressure at 60°C:

$$s\ (g/L) = 0.0249\ P_{ClO2}\ (mmHg)$$

Based on the theoretical amount of ClO_2 generated (132 mmol), in the worst case, there will be in the head space over the solution about 50% of chlorine dioxide and thus an explosive atmosphere.

Therefore, we tried to modify the conditions of delignification to maintain ClO_2 concentration in the gas phase below 10%, i.e. its safety limit (Table 3.1). At the same time, we wanted to maintain the same delignification efficiency which can be measured by the loss of beech wood weight after this treatment.

Exp	AcOH (mmol)	NaClO₂ (mmol)	V H₂O vs beech wood	Time (h)	T (°C)	Beechwood weight loss (%)	ClO₂ in the atmosphere
1	3x44	3x118	10	3	60	10.5	50%
2	14	18	10	24	60	5.7	<10%
3	28	36	20	24	60	7.0	<10%
4	49	62	10	24	25	6.6	<10%
5	98	124	20	24	25	6.6	<10%
6	-	-	10	24	25	3.9	-

Table 3.1. Modification of ClO_2 delignification conditions.

First of all, we tried to carry out the reaction with amounts of acetic acid and sodium chlorite which generate less than 10 % of ClO_2 in the atmosphere, maintaining 60°C as the reaction temperature and using 10 V of water *vs* beech wood (exp 2). In the third experiment we doubled the amount of water to replicate the absolute amount of chlorine dioxide, but keeping constant the final concentration. In experiments 4 and 5 we decreased the reaction temperature to 25°C in order to increase ClO_2 solubility in water. Indeed, according to IUPAC solubilities data[21], the relationship between solubility in water and partial pressure at 25°C is the following:

$$s \ (g/L) = 0.0084 \ P_{ClO2} \ (mmHg)$$

However, in all the trials (exp 2-5) the beechwood weight loss was not more than 7%, which is lower than 10.5 % obtained in the first experiment. Moreover, you can observe that the use of neat water (without any oxidant agent) causes a weight loss of 3.9% indicating that part of the weight loss is due to the so-called extractives (e.g. phenols, terpenes). Therefore, the amount of lignin removed in experiments 2-5 is low (below 3%) and does not allow xylan extraction in the following step. From these preliminary experiments we concluded that it would be difficult to have chlorine dioxide delignification inherently safe (ClO_2 gas < 10%) and concomitantly efficient.

For lab trials we use the conditions suggested by Dr. Capek but using a nitrogen flow to push the excess of chlorine dioxide towards two scrubbers working with sodium thiosulfate, which is reported to react quickly with ClO_2.[22]

3.4b Xylan extraction from ClO$_2$ delignified sawdust

Once delignified, beech wood sawdust can be extracted by two methods. The first one uses dimethylsulfoxide (DMSO) as reported in literature.[23,24] The extraction can be carried out at different temperatures (25-70°C) and is usually repeated twice. At the end of the extraction the xylan is isolated by precipitation in ethanol and acidification with HCl. We applied this procedure performing a first extraction at 60 °C for 24 h and then we tried to isolate the product as described above. However, we obtained a solid hard to filter and the recovery was very low (yield: weight of isolated product/weight of biomass = 0.25%). Moreover, the product is probably contaminated by a high amount of salts and solvent because in GPC-RI a high peak at low MW (<1000 Da) is observed and no other peak at higher MW is detected. On the other hand, it was possible to record ^{1}H-NMR spectrum, which despite the low resolution shows a series of signals between 5.5 and 3.0 ppm related to protons of the xylose chain (Figure 3.7).
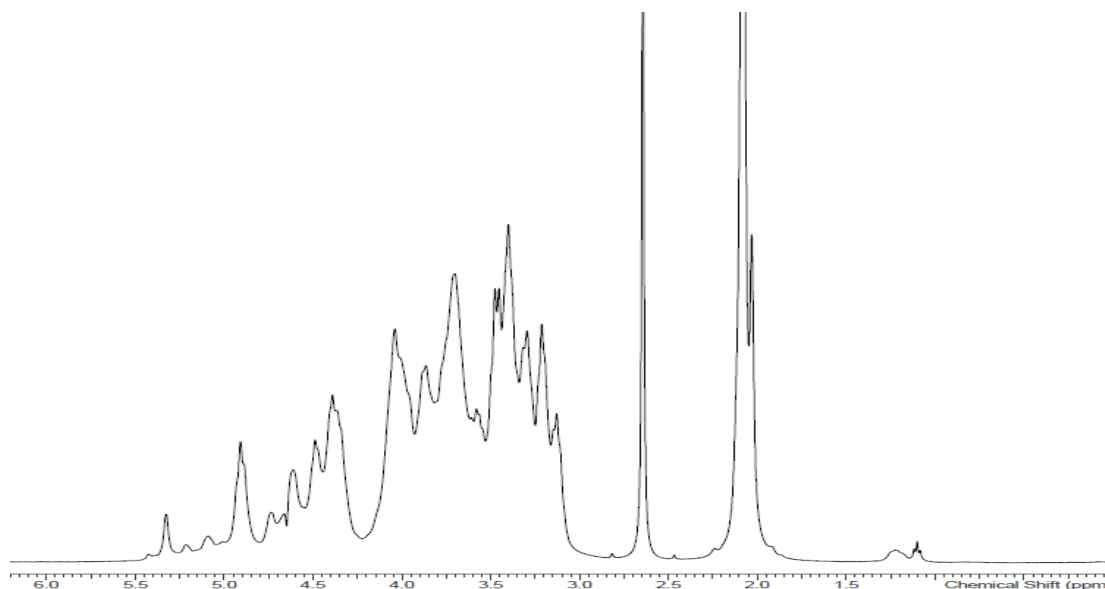
Figure 3.7. ^{1}H-NMR spectrum (D$_2$O) of xylan extracted by DMSO

It is also possible to notice an intense signal at 2.6 ppm related to residual DMSO and another one at 2.1 ppm linked to the presence of acetyl groups. The low resolution of the spectrum does not allow to differentiate and identify accurately the position of the acetyl groups, but the intensity of the signals suggests that almost all the acetates of the chain were kept intact and hence the xylan has acetyl content and profile different from Originator SP.

An alternative method is the use of aqueous alkaline solutions, which were widely used in the paper industry and allowed to obtain good extraction yields.[25] The main issue is associated to the complete *O*-deacetylation caused by the basic conditions.[6] Therefore, in collaboration with Dr. Capek (Slovak Academy of Science) we developed a mild alkaline method of extraction which

extracts the xylan but keeps intact a portion of the acetyl groups. According to this procedure, the delignified sawdust is extracted twice with 0.5-1% NH_3 at room temperature and then the xylan is isolated by precipitation in EtOH. Compared to the DMSO method, the recovery (weight of isolated products/ weight of delignified sawdust) is higher: y (DMSO) = 0.25% *vs* y (NH_3) = 2.24%. Moreover, the ^{1}H-NMR spectrum (Figure 3.8) of the ammonia extract revealed a different acetyl profile compared to DMSO.
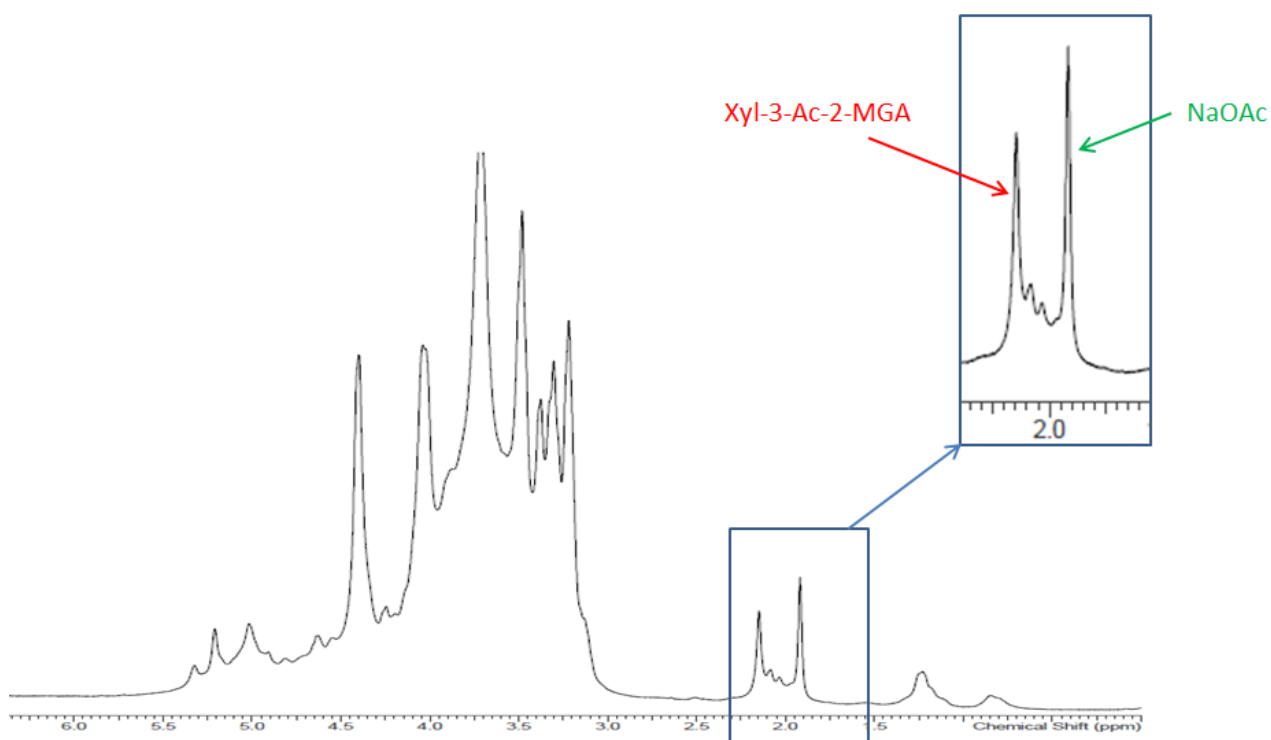
Figure 3.8. ^{1}H-NMR spectrum of xylan extracted by NH_3. In the blue square there is a zoom of acetates region.

First of all, it is evident that acetyl signals are less intense and hence a partial deacetylation of the xylan occurred during the extraction in slightly basic conditions. We can observe a signal at 1.85 ppm due to the presence of free acetate. This was confirmed by a spike experiment with sodium acetate. The other intense singlet at 2.1 ppm is the acetyl group in position 3 of a xylose bearing MGA residue[26], which is the most abundant acetyl present in the Originator SP. On the other hand, the less intense singlets between 2.1 and 1.9 ppm are related to acetyl groups in position 2 or/and 3 of other xylose units. Probably the basic conditions hydrolyze almost all the acetates of the chain, but the one present in position 3 of xylose bearing MGA is kept preferentially because of steric hindrance. From ^{1}H-NMR analysis we can conclude that the extraction with diluted ammonia solution allows to obtain a profile in the ^{1}H-NMR acetyl area very similar to the Originator SP.

We analyzed the glucuronoxylan with GPC-RI and Mn was 2956 Da and Mw 12426 Da with a high polydispersity of 4.2. The molecular weight of the polymer is too high to reach the desired MW distribution of the Originator SP and hence it is necessary to depolymerize chemically the xylan (paragraph 3.6).

3.5 Xylan extraction: Peracetic acid method

The second method is similar to the first one, because it is characterized by delignification of beech wood sawdust before extraction of the xylan. However, in this case the delignification step is carried out with peroxyacetic acid instead of chlorine dioxide.

3.5a Beech wood delignification with peracetic acid

Peroxyacetic acid has an oxidizing power similar to chlorine dioxide and hence has a similar delignifying ability and generates similar products (Scheme 3.3).[27]
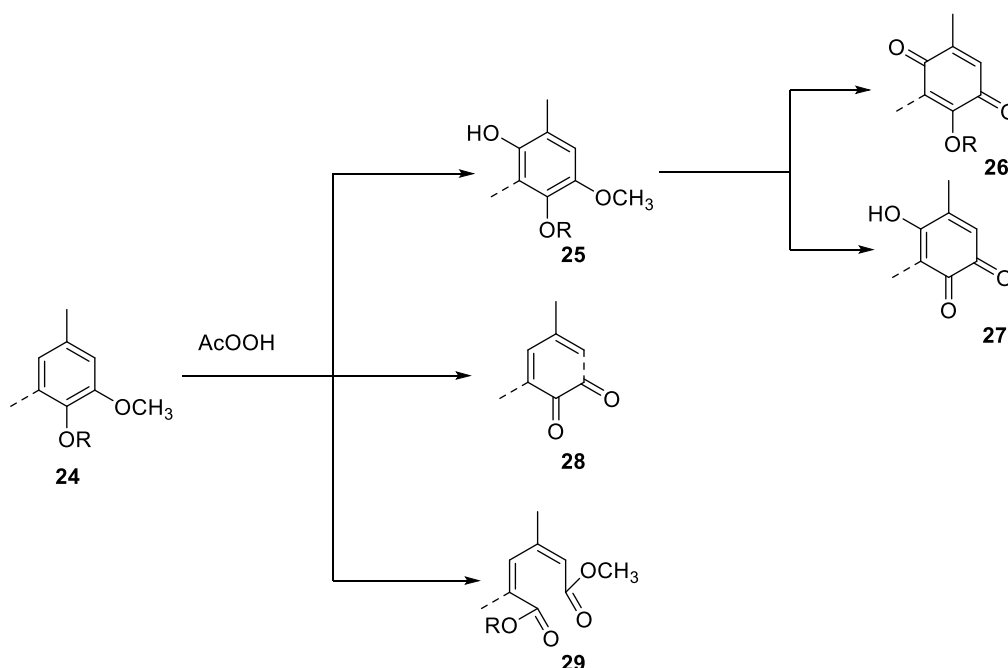
Scheme 3.3. Oxidation of aromatic structures of lignin by peracetic acid. R = H or Me.

The reactive species during lignin oxidation is the hydroxonium ion, HO^+, which is generated by the heterolytic cleavage of the peroxidic bond.[17] Hydroxonium ions can attack the aromatic ring **24** in *para* position to the methoxy group, generating the hydroxylated species **25**, which can be further oxidized to 1,4-quinone **26** or 1,2-quinone **27**. The aromatic structure of lignin **24** can also be directly oxidized to quinone **28** or the aromatic ring can undergo oxidative cleavage to muconic acid derivative **29**.

Moreover, the hydroxonium ion from peracetic acid is also able to cleave β-aryl ether bonds, which are the main linkage present in lignin structure (Scheme 3.4).[28] The β-aryl ether portion of lignin **30** can evolve to diol **31** and quinones **32** via two possible mechanisms which involve the attack of HO⁺ to hydroxylated positions of the aromatic ring in blue and a subsequent reaction with water. Structure **30** can also generate α-hydroxy-ketone **33** and creosol **34**.
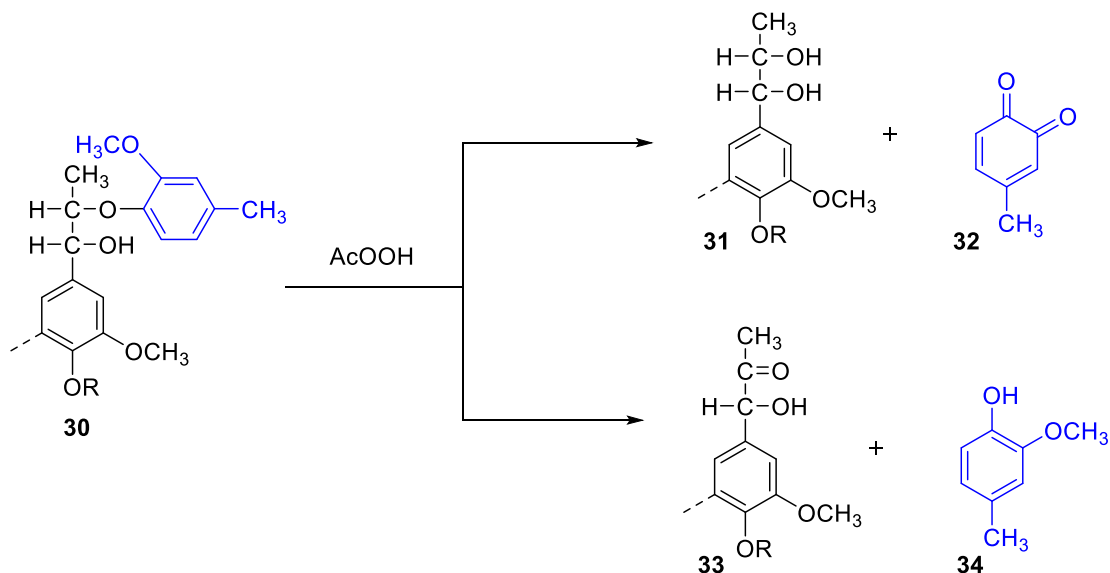
Scheme 3.4. Cleavage of β-aryl ether structures by peracetic acid

The first trial of delignification was performed with the conditions reported in the literature[29]: sawdust was treated with 40 volumes of 11% w/w AcOOH solution at 85°C for 25 min. The weight loss of the sawdust after this delignification was 38%, which was much higher than the value obtained with chlorine dioxide (10.5%). Moreover, because in beech wood the lignin content is about 28%, it is likely that peracetic acid removed not only lignin, but also small portions of cellulose and glucuronoxylan. The main drawback of this procedure is the use of 40 volumes of 11% peracetic acid solution *vs* beech wood sawdust and this means low productivity. Therefore, we ran a few trials to decrease the volumes of the solution, possibly without losing efficiency in the removal of lignin (Table 3.2). Entry 1 shows the results obtained with the literature conditions described above. In the second experiment (Entry 2) we decreased the volumes from 40 to 8 while keeping all the other parameters constant and a satisfying weight loss (30.7%) was obtained. However, we observed a strong frothing in the mixture, due to fast and abundant release of oxygen, which makes unsuitable the use of these conditions for higher scale. Therefore, to slow down the reaction kinetics and to keep under control the release of oxygen, we carried out the delignification at lower temperature (room temperature in experiment 3 and 50°C in experiment 4), but the weight loss decreased significantly (8.1% and 24.3%, respectively). To compensate the

lower temperatures, we increased peracetic acid concentration to 22% and at 50°C (Entry 6) we obtained a good weight loss (27.5%). However, it seems that achieving a weight loss above 30% requires to raise the temperature to 85°C and thus in experiment 7 we went back to the conditions used in Exp. 2, but modified the mode of AcOOH addition. Indeed, in all the experiments from 1 to 6 an 11% w/w AcOOH solution was prepared and then the sawdust added. In experiment 7 the sawdust was suspended in 5.5 volumes of water at 85°C and 2.5 volumes of 40% AcOOH were added drop-wise. The final concentration of peracetic acid is 11% and the total volumes 8, but the dosage allows controlling the release of oxygen. In these conditions we obtained a weight loss (38.0%) comparable to experiment 1.

Exp.	AcOOH conc. (w/w)	Volumes of AcOOH solution vs beech wood	t (h)	T (°C)	% Beech wood weight loss	Notes
1	11%	40	0.5	85	38.8	
2	11%	8	0.5	85	30.7	Strong bubbling
3	11%	8	20	r.t.	8.1	
4	11%	8	4	50	24.3	
5	22%	8	20	r.t.	20.7	
6	22%	8	4	50	27.5	
7	11%	8	1 + 0.5	85°C	38.0	

Table 3.2. Investigation of conditions for peracetic acid delignification. All the experiments were run on 15 g of sawdust.

The conditions of experiment 7 seem the most promising ones, because the release of oxygen is under control and the high weight loss should determine a high extraction yield in the following step. Therefore, we decide to use these conditions to scale up the delignification from 15 g to 242 g of sawdust and we obtained 128 g of delignified sawdust (weight loss = 42.4%). However, to control the release of oxygen on a larger scale the addition of peracetic acid was slowed down from 1h to 3h.

3.5b Xylan extraction from AcOOH delignified sawdust

After delignification it is possible to extract the xylan and, as established in the chlorine dioxide experiments, we used diluted ammonia at room temperature. The xylan was isolated by precipitation with EtOH and analyzed by ^{1}H-NMR (Figure 3.9). Comparison of the spectra of the

xylan obtained after the two delignification methods indicates a high similarity. The acetyl groups profile at about 2 ppm is very similar and thus we can deduce that ammonia extraction allows to obtain the desired acetyl content independently from the delignification method.
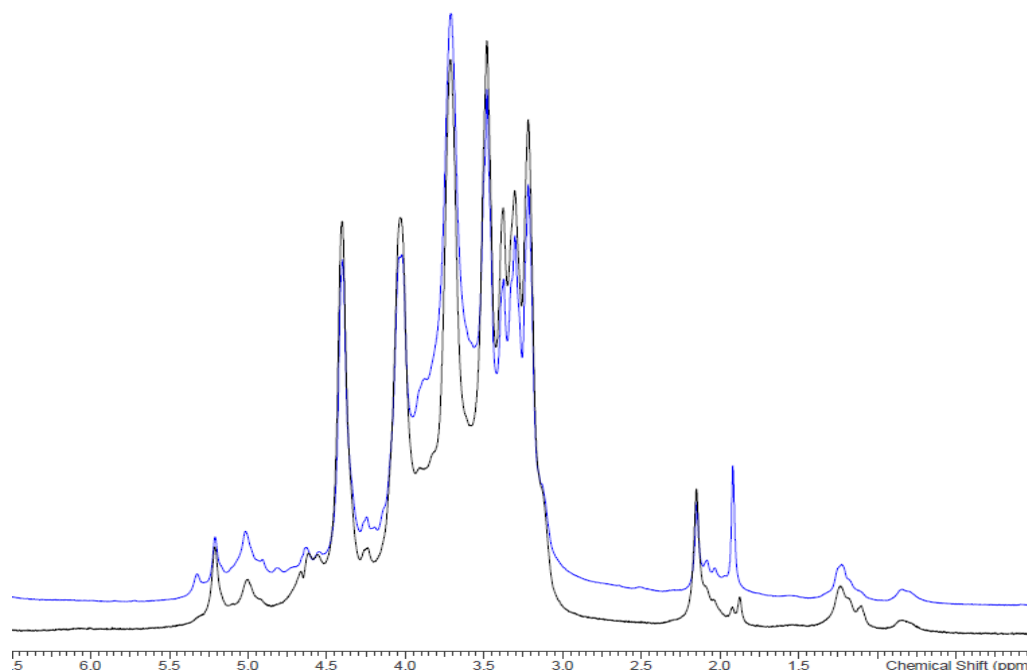
Figure 3.9. ^{1}H-NMR spectra of xylans obtained after chlorine dioxide (blue) and peracetic acid (black) delignification.

NMR analysis suggests that the delignification method does not affect the quality of the glucuronoxylan. On the other hand, GPC revealed an important impact of the delignification method on the molecular weight distribution. In Figure 3.10 the chromatograms of xylan obtained after chlorine dioxide delignification (top) and with peracetic acid method (bottom) are shown. The ClO$_2$ chromatogram reveals two main peaks: the first one is the xylan and the second one is related to the presence of inorganic salts (e.g. chlorides). The AcOOH chromatogram shows an additional peak at lower retention time i.e. higher molecular weight. This peak can be caused by the formation of aggregates or by contamination of cellulose which was partially extracted. To clarify this feature, we sulfated the xylan with the procedure which will be discussed in detail in paragraph 3.8 and we observed the disappearance of the third peak in GPC. Therefore, we can conclude that cellulose was not extracted, but the xylan can form aggregates (perhaps independently from the delignification method) which are destroyed by the sulfation process.
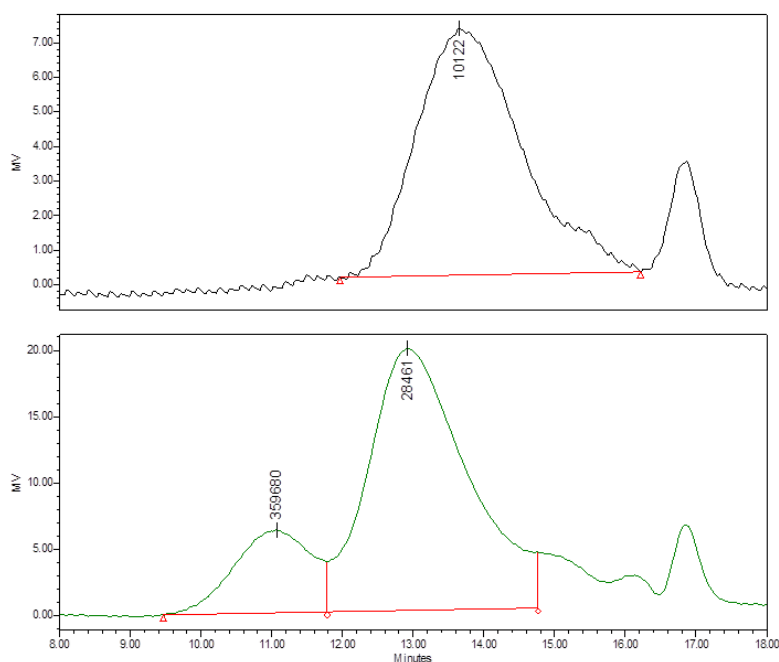
Figure 3.10. GPC chromatograms of xylans after chlorine dioxide (on the top) and peracetic acid (on the bottom) delignification.

Moreover, GPC shows different molecular weight values of the glucuronoxylan obtained with the two methods, as reported in Table 3.3. The molecular weight parameters of xylan after peracetic acid delignification are higher, almost three times the values obtained with the chlorine dioxide method. AcOOH delignification is more extensive than the one carried out with ClO_2, as suggested by the different weight loss (42% vs 10.5%), and hence it allows the extraction of high molecular weight oligomers, which are otherwise trapped by lignin. To support this hypothesis, the extraction yield from AcOOH delignified sawdust is also higher (4.0% vs 2.2%).

Exp	Delignification Method	Mn	Mw	MP	PD
1	ClO_2	2956	12426	11292	4.2
2	AcOOH	9583	28470	28461	3.0

Table 3.3. GPC results of xylans extracted after different delignifications.

In conclusion, among the two delignification methods peracetic acid seems the most promising because it has less safety issues and allows a higher extraction yield. However, in both cases the molecular weight distribution of the xylan is not suitable to reach the target after sulfation and the product needs to be depolymerized to molecular weight of about 1500-2500 Da.
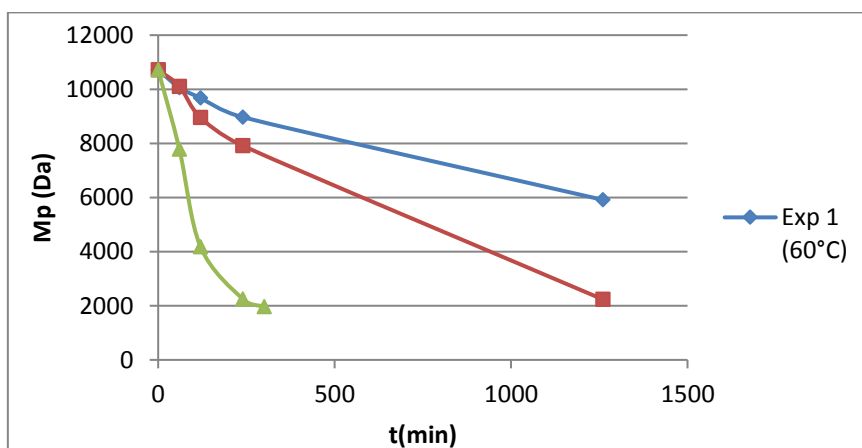
3.6 Xylan depolymerization

The depolymerization of the glucuronoxylan can be achieved essentially *via* enzymatic or chemical pathways. The enzymatic degradation involves the use of xylanases, a class of enzymes able to hydrolyze the glycosidic linkage between xylose units. Xylanases are usually endo-type (they cut the polymer in its middle) and attack the xylan in a random manner allowing a significant decrease in the molecular weight.[30] Literature reports examples of depolymerization of Fagus glucuronoxylan by these enzymes.[31] The other depolymerization strategy is based on the use of chemical reagents and among them the most interesting ones are based on hydrogen peroxide. Hydrogen peroxide can be activated by Fe^{2+}, which promotes the formation of hydroxyl radicals ($OH^{.}$). These radicals can attack the glycosidic bond yielding a heterolytic cleavage with formation of a radical center on the xylan chain which propagates the reaction and the degradation of the xylan to lower molecular weight fragments.[32] Moreover, hydrogen peroxide can also be used in combination with sulfuric acid to carry out the so called oxidative-acidic depolymerization.[33] This method is described to retain the glucuronic acid content and to generate oxidation species like carbonyl groups and double bonds conjugated systems, which are present in the originator sulfated polysaccharide. In several patents this method is exploited but it is usually carried out on the sulfated glucuronoxylan probably because of its higher solubility in water.[34,35] However, the oxidative-acidic depolymerization causes partial removal of the sulfates and hence it would require re-sulfating the product. To avoid this drawback, it would be more practical to depolymerize the glucuronoxylan before sulfation, which should be possible despite the modest solubility in water. To investigate this possible route, we ran preliminary experiments on commercially available xylans (GX). They are not suitable for the preparation of the sulfated polysaccharide as discussed in paragraph 3.3, but they are very similar to extracted partially acetylated xylans (AcGX). Therefore, they can be used as model compounds to study the kinetic and the conditions of the reaction without scarifying the xylans extracted from beech wood in the lab.

The oxidative-acidic depolymerization was carried out with 30% H_2O_2 (0.5 mL/g xylan), 2.5 M H_2SO_4 (0.04 mL/g xylan) and water (3V). To monitor the kinetics, an aliquot of the reaction mixture was neutralized with 0.1 M NaOH and after removal of the inorganic sulfates by precipitation with 1M $BaCl_2$ solution, the sample was analyzed by GPC-RI. GPC chromatograms of these IPC (in-process controls) are perturbed by the solvent and an accurate evaluation of the molecular weight distribution is not possible. However, the apex of the peak, Mp, is not very affected by the quality

of the chromatogram and hence it is a significant indication of the molecular weight of the polymer. We ran three experiments on the Senn chemicals glucuronoxylan (GX) maintaining constant the amount of reagents and solvent (Table 3.4). We explored three different temperatures (60, 80 and 100°C) and, at the end of the experiments, we isolated the glucuronoxylans by precipitation in Methanol. First of all, in all the trials Mp of the last IPC (after 1260 min for Exp.1 and Exp2, after 300 min for Exp.3) is lower than Mp of the isolated xylan. It is likely that the low molecular weight fragments are more soluble and are lost in the mother liquor of the final precipitation, determining an overall increase of the molecular weight distribution. Therefore, it is clear that we need to have Mp below 2000 Da in the last IPC to reach the molecular weight target of the xylan after isolation (1500-2500 Da). Moreover, as we can clearly see in Figure 3.11, the temperature plays an important role in the reaction rate showing that the depolymerization is too slow at 60 and 80°C. On the other hand, at 100°C (Exp.3) after 5 h Mp is close to the desired value.

	Mp (Da)		
t (min)	Exp. 1 (60°C)	Exp. 2 (80°C)	Exp. 3 (100°C)
0	10721	10721	10721
60	10060	10106	7790
120	9679	8950	4178
240	8967	7911	2249
300	-	-	1969
1260	5915	2236	-
isolated	6793	3426	2818

Table 3.4. Kinetic study of oxidative-acidic depolymerization of xylan (GX).

Figure 3.11. Graphical representation of kinetic study of oxidative-acidic depolymerization of xylan (GX).

According to these experiments, the oxidative-acidic depolymerization at 100°C after 5-6 h allows to reach the desired molecular weight. Therefore, we applied these conditions to an acetylated xylan (AcGX) extracted from beech wood with AcOOH method (Mp = 17 KDa). However, the conditions were not directly transferable as can be seen Figure 3.12. Indeed, the molecular weight of the acetylated xylan was reduced below 1 KDa after only 2 h. The reason of this difference is not clear and is hardly related to the presence of a small portion of acetylated hydroxyl group. It might be more plausible that our extraction method causes contamination of the xylan with an impurity (e.g. metals), which significantly accelerates the reaction rate.
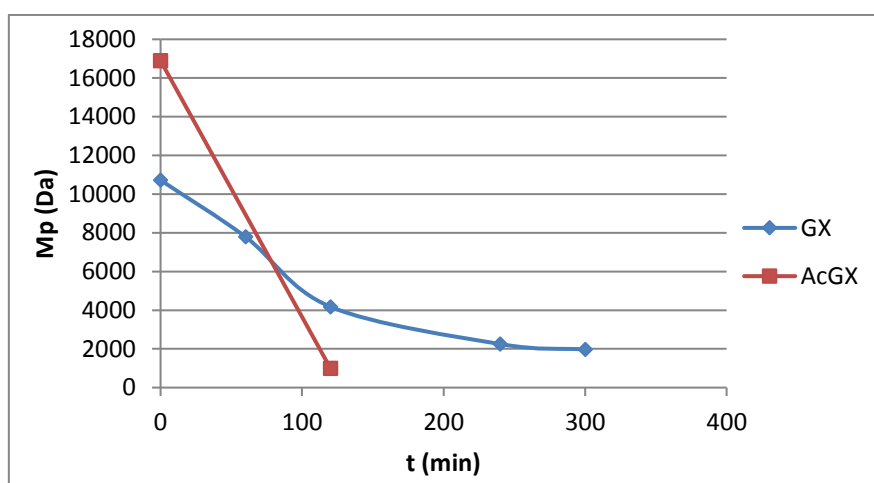
Figure 3.12. Comparison of depolymerization at 100°C of commercially available (GX) and acetylated xylan (AcGX).

Anyway, it is clear that the oxidative depolymerization of the acetylated glucuronoxylan (AcGX) has to be carried out at lower temperature. We ran an experiment at 70-75°C and we observed a controlled decrease of the molecular weight (Figure 3.13).
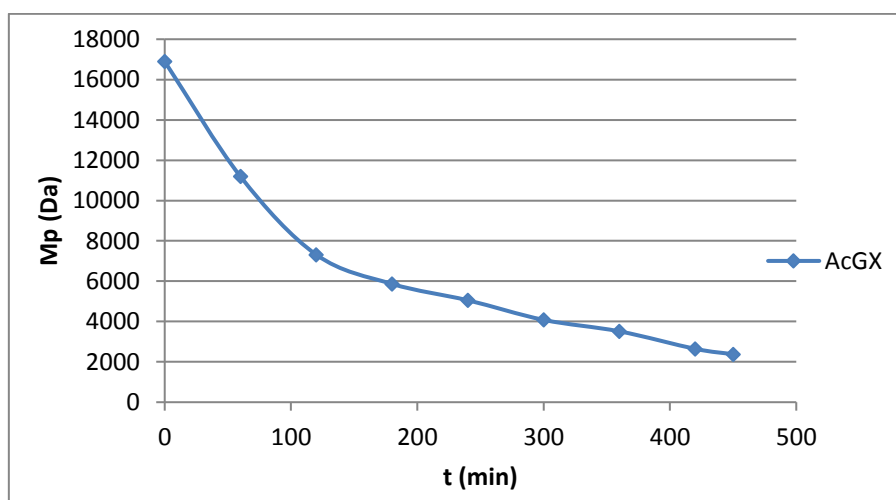
Figure 3.13. Depolymerization of acetylated glucuronoxylan at 70-75°C.

According to IPC, after 8 h at 70-75°C Mp of the xylan is about 2 KDa, so we quenched the reaction mixture and we isolated the xylan with Mp of about 2.5 KDa.

We analyzed the obtained glucuronoxylan also by NMR finding that the spectra before (blue spectrum) and after (red spectrum) depolymerization are very similar (Figure 3.14). The signal at about 5.2 ppm, the anomeric proton of 4-*O*-methylglucuronic acid, and the group of signals at about 2 ppm, the acetyl groups, are at a first glance unchanged. Therefore, ^{1}H-NMR spectrum seems to confirm that there is no modification of MGA and acetyl content during depolymerization, as claimed in literature.
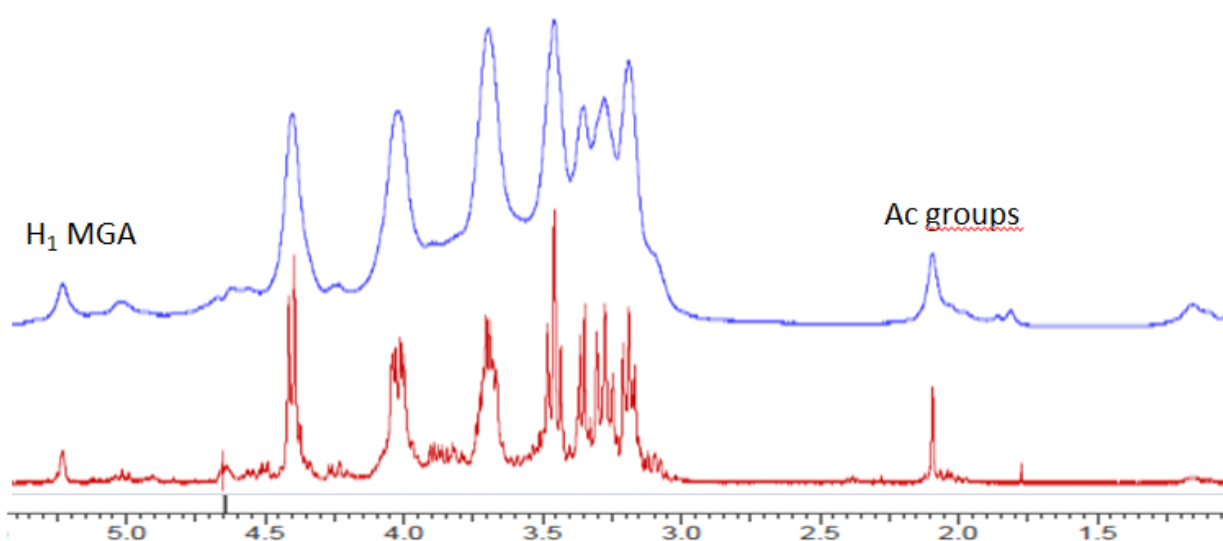
Figure 3.14. Comparison of ^{1}H-NMR spectra of acetylated xylans (AcGX) before (blue) and after (red) depolymerization.

In summary, we have found the proper conditions to depolymerize the acetylated glucuronoxylans obtained from beech wood and to obtain a molecular weight similar to Originator SP after sulfation. The depolymerization at a first glance does not seem to affect the content of acetyl groups and 4-*O*-methylglucuronic acid. This last aspect would anyway need to be further investigated.

3.7 Xylan extraction: autohydrolysis

The two methods discussed until now are characterized by three main steps: delignification (chlorine dioxide or peracetic acid), ammonia extraction and oxidative-acidic depolymerization. The third strategy has a completely different sequence and is known in literature as autohydrolysis.[36] The procedure is very simple because sawdust is suspended in water at high

temperatures (150-200°C). It is similar, but milder than steam explosion, the treatment of biomasses with superheated steam at 20-40 atm for a short time (30-300 s) followed by a fast decompression.[37] The autohydrolysis is able to degrade and extract hemicellulose leaving intact lignin and cellulose. The name is linked to its mechanism of action, which is an auto-catalyzed hydrolysis of the xylan. Indeed, at high temperature water cleaves and extracts acidic compounds: a portion of the acetyl groups are hydrolyzed to acetic acid, the uronic and phenolic acids are liberated in solution. All these species decrease the solution pH, which accelerates the depolymerization and the extraction of the glucuronoxylan. The exact degradation mechanism is not clear, but it is plausible that the xylan is degraded to low molecular weight fragments which are easily solubilized in water. Alternatively, the xylan is first extracted as high molecular weight oligomers and then depolymerized. Anyway, at the end of the treatment the liquor contains the xylan with a low degree of polymerization, xylose residues and oxidation products such as furfural or hydroxymethylfurfural. As far as the acetyl groups are concerned, it has been described that the deacetylation rate is similar to the rate of xylan extraction.[38] It seems hence possible to extract a partially acetylated xylan with this method, but temperature and time of autohydrolysis are crucial to control the extent of the deacetylation.

The main advantage of this approach is that the glucuronoxylan can be extracted without any prior delignification and this allows to avoid the use of hazardous reagents difficult to handle industrially. On the other hand, the autohydrolysis needs high temperature and consequently high pressures, which can be source of issues in an industrial plant. Therefore, we started working on autohydrolysis by evaluating if it was possible to extract the xylan with reasonable yields at temperatures equal or below 150°C, a range below the values described in literature (Table 3.5). In all the trials at the end of the autohydrolysis the xylan was isolated by concentration of the mother liquor and precipitation in methanol. The first experiment was carried out at 150°C for 1h and we observed good extraction and isolation yields of the xylan (3.5%). Decreasing the temperature to 120°C requires longer reaction time because after 3h (Exp. 2) the dry residue is very low (3.3%) and it is not possible to isolate any material. After 21 h (Exp 3) the dry residue is high (15.9%) and the xylan is isolated with 6% yield. In the last experiment we tried to further decrease the temperature in order to avoid any overpressure, but at 100°C the autohydrolysis does not start and after 48 h the dry residue is below 3 % and pH is still 5. This means that the concentration of hydronium ions at 100°C remains low because the acidic components are not liberated in solution and hence the xylan is not extracted.

163

Exp	T (°C)	P (barg)	Time (h)	Final pH	Dry residue vs wood (%)	Yield of isolated xylan vs wood (%)
1	150	3.8	1	4.1	9.0	3.5
2	120	1.0	3	4.6	3.3	-
3	120	1.0	21	3.8	15.9	6.0
4	100	0	48	5.0	2.9	-

**Table 3.5. Preliminary evaluation of temperature and time of autohydrolysis.**

From these preliminary experiments we concluded that the glucuronoxylan can be extracted with good yields (higher than with the ClO_2 and AcOOH methods) between 120 and 150°C causing an overpressure of 1 or 3.8 barg respectively, a range of temperatures and pressures easily achievable in an industrial plant.

The isolated xylans were analyzed by GPC-RI and the Mp parameter was about 3000-3500 Da, significantly lower than the molecular weight of xylans obtained with the previous extraction methods. These results confirm that with autohydrolysis, as described before, the hydronium ions catalyze the depolymerization to low molecular weight oligomers. To better understand the course of autohydrolysis, we ran a few trials sampling the mixture at different times and studying the variations in molecular weight, dry residue and pH. For example, in Table 3.6 and Figure 3.15 the results of a trial performed at 130°C (P =1.7 barg) for 19 h are shown. A decrease of pH was observed (Figure 3.15, blue) because acidic species (acetic and uronic acids) were progressively extracted. Moreover, the dry residue (Figure 3.15, red) raised from 2 % to 14 % with almost a linear trend ($R^2 = 0.98$). GPC-RI analysis could not detect any peak in the first two samples because the concentration of the xylan in solution was too low, while from 9 h to 19 (Figure 3.15, green) it revealed a decrease of Mp, the apex of the peak, from 2 KDa to 0.8 KDa.

t (h)	pH	Dry residue vs Wood (%)	Mp (KDa)
0	5.54	2.32	n.d.
4	4.33	3.67	n.d.
9	4.11	8.95	2.024
14	3.95	11.35	1.118
19	3.87	14.28	0.808

Table 3.6. Autohydrolysis course at 130°C (P =1.7 barg). n.d. = not detectable.
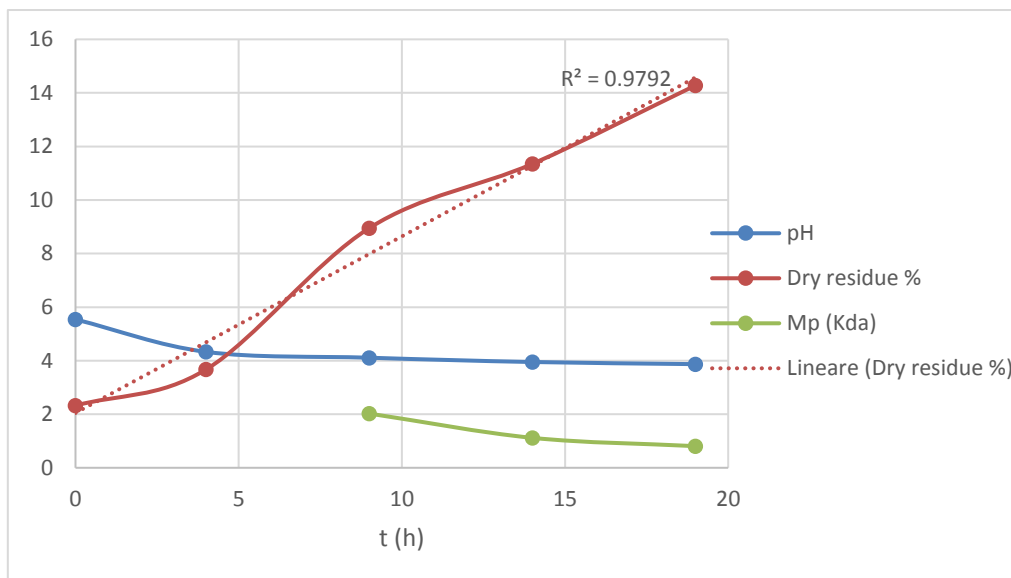
**Figure 3.15. Autohydrolysis course at 130°C (P =1.7 barg).**

The data show that autohydrolysis depolymerize the xylan to low molecular weight fragments and hence with this method the depolymerization occurs simultaneously to the extraction. It is not necessary to introduce an additional step of depolymerization as in the other two methods. The choice of temperature and time of autohydrolysis is crucial to obtain the desired molecular weight distribution. The data reported are referred to samples of mother liquors from which the xylan must be isolated by precipitation. As discussed in the last paragraph (oxidative acidic depolymerization) during precipitation a portion of oligomers with low degree of polymerization is lost in the mother liquor and hence the MW distribution of the xylan changes. Therefore, if we want to isolate a xylan of 2 KDa, we have to stop autohydrolysis to about 1 KDa of Mp. Several trials were carried out with different autohydrolysis and xylan isolation conditions in order to obtain the desired molecular weight distribution. They are not reported here for confidentiality reasons, but we did manage to obtain satisfying results.

The xylan obtained by autohydrolysis was also analyzed by ^{1}H-NMR and the spectrum is reported in black in Figure 3.16. The comparison with the spectrum of a xylan obtained by ammonia extraction (blue in Figure 3.16) reveals a relevant difference in the acetyl groups signals at about 2 ppm. Specifically, the xylan obtained with ammonia has an intense singlet of acetyls in position 3 of a xylose bearing an MGA unit and the other signals have very low intensity. On the other hand, the xylan coming from autohydrolysis has a higher acetates content and a different signals pattern, indicating acetyl groups located in other positions of the xylose chains.
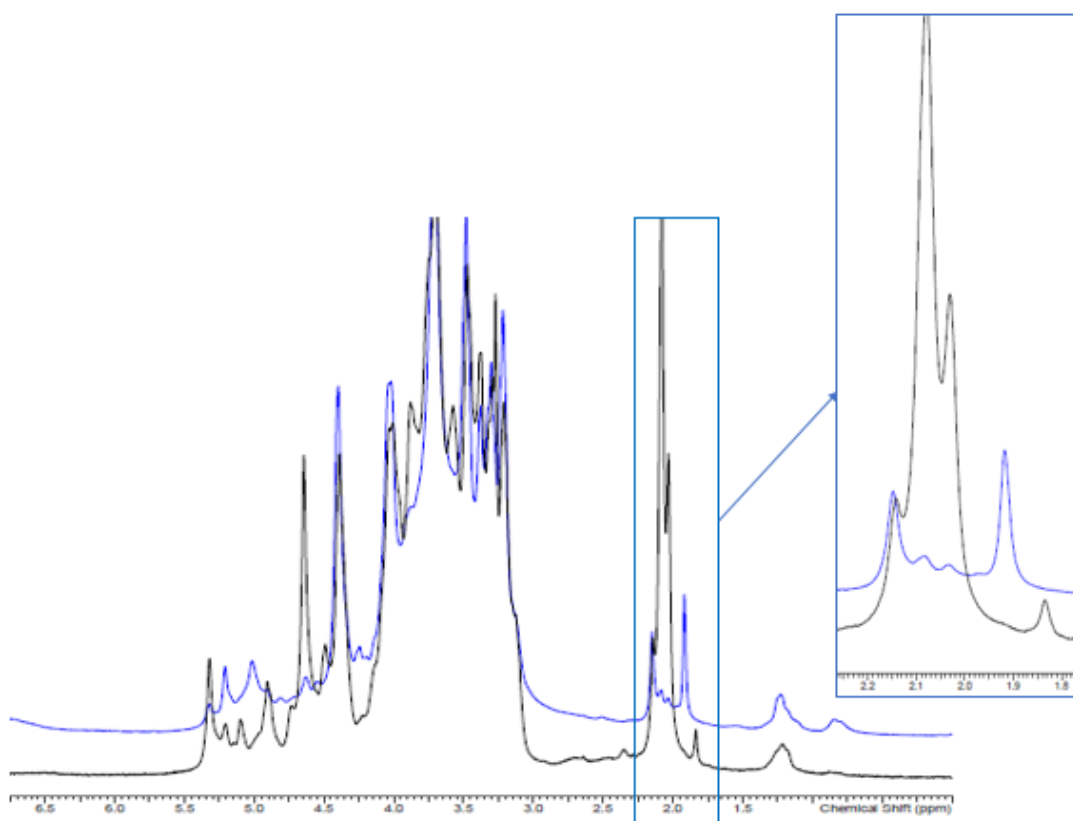
Figure 3.16. Overlay of ^{1}H-NMR spectra of xylans obtained by ammonia extraction (blue) and autohydrolysis (black).

It is clear from ^{1}H-NMR analysis that the autohydrolysis xylan is over acetylated relative to Originator SP and needs to be selectively deacetylated. We ran a preliminary deacetylation experiment by dissolving the xylan in deuterated water and adding a small amount of a base (below 1% w/w). We monitored the behavior of the acetyl groups in these conditions by ^{1}H-NMR (Figure 3.17). First of all, the NMR spectra at different times are not exactly aligned because the solution pH changes during the deacetylation and causes small variation in the chemical shifts. However, a strong increase of the signal at 1.85 ppm was observed and it indicates that the acetyl groups are removed from the chain and are liberated in solution as free acetates. Moreover, among the three singlets present at the beginning, the signal at about 2.18 ppm decreases much more slowly than the other ones and is still present after 23 h. Reasonably, these acetyl groups in position 3 of a xylose bearing MGA in 2 are less prone to hydrolysis because of steric hindrance. From this study we can conclude that under mild basic conditions it is possible to selectively deacetylate the autohydrolysis xylan maintaining mainly the acetyl groups in the desired position. To obtain a similar acetates pattern of the Originator SP it is important to control the concentration of the base and the reaction time (these experiments are not reported here for confidentiality reason).
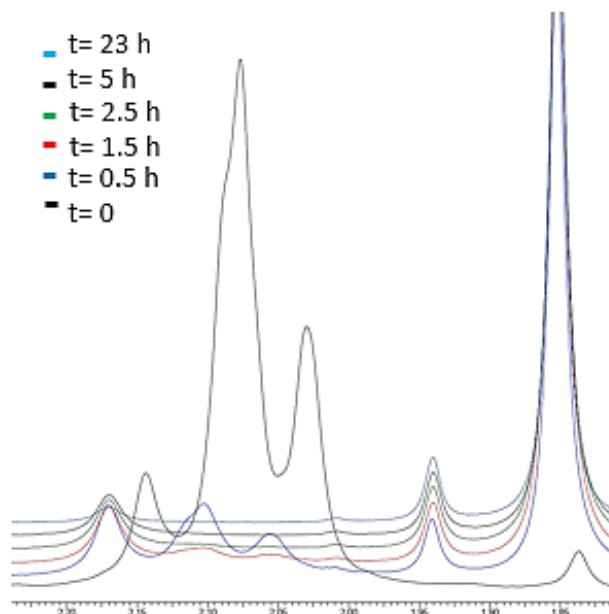
Figure 3.17. ^{1}H-NMR monitoring of deacetylation of autohydrolysis xylan.

Thus, the process was developed in order to avoid the isolation of the over-acetylated xylan, as shown in Scheme 3.5. The green process is characterized by isolation only of the partially acetylated xylan. Specifically, at the end of the autohydrolysis the mother liquors are concentrated, deacetylated, further concentrated and precipitated with methanol. Avoiding an isolation step is advantageous for an industrial and economic perspective and allows not to lose low molecular weight oligomers in the precipitation mother liquors thus avoiding to increase excessively the molecular weight of the final glucuronoxylan
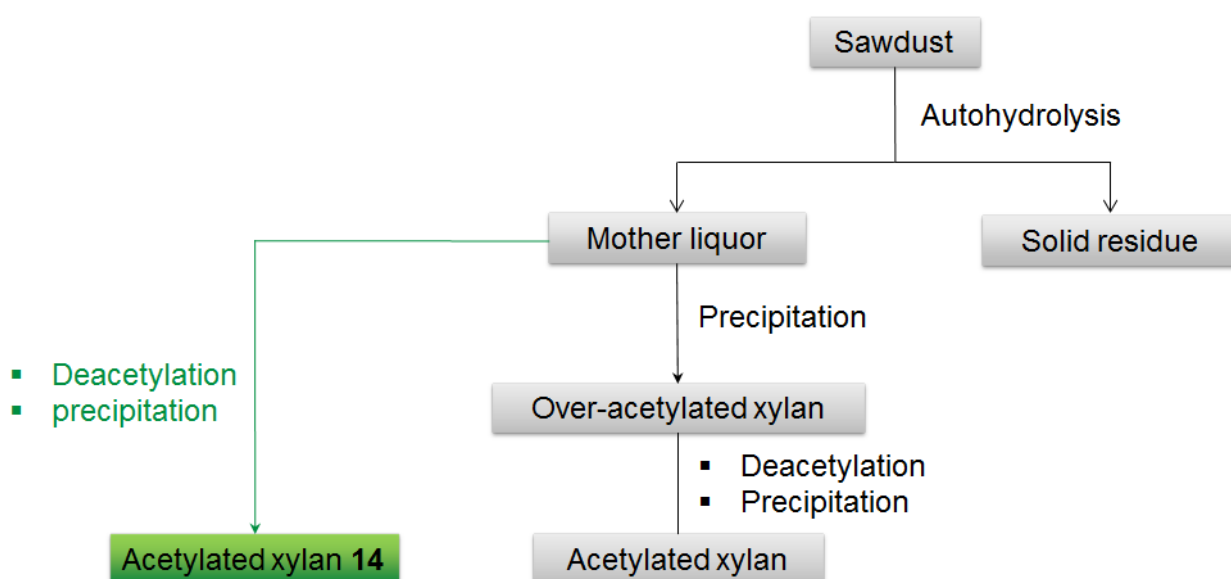

Scheme 3.5. Flow diagrams of the isolation of a partially acetylated xylan. The green process avoids the isolation of the over-acetylated xylan.

167

In conclusion, the autohydrolysis and the following deacetylation process appear as the most promising strategy to obtain the glucuronoxylan for preparation of the SP. Indeed, this method does not need any prior delignification and hence does not use hazardous chemical reagents. Moreover, the autohydrolysis mechanism allows to depolymerize the xylan simultaneously to its extraction, thus avoiding the introduction of an additional step of chemical depolymerization. Therefore, we decided to scale-up this process on pilot plant obtaining about 1 Kg of partially acetylated glucuronoxylan from 25 Kg of beech wood sawdust. The xylans obtained with this method are the starting materials for all the experiments which will be discussed in next paragraphs.

3.8 Sulfation of the xylan

Xylan **14** has to be sulfated to yield the sulfated polysaccharide (SP) **1** (Scheme 3.6). The reaction is carried out at 60-80°C with sulfur trioxide-pyridine complex **35**, which has been routinely used for the sulfation of carbohydrates. The solvent can be pyridine itself [34] or a mixture of pyridine and *N,N*-dimethylformamide (DMF).[35,39] We focused on the use of pyridine as solvent to avoid the introduction of an additional solvent hard to remove from the final product.
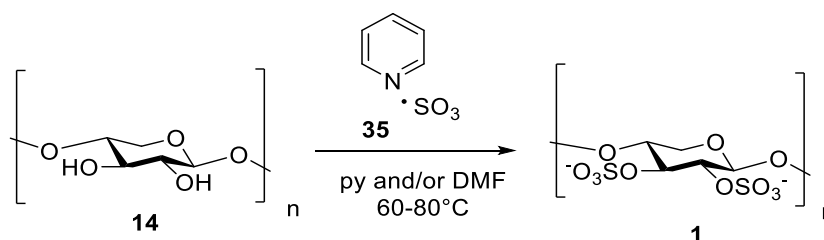
Scheme 3.6. Sulfation of the glucuronoxylan.

Sulfur trioxide is an electron acceptor or Lewis acid and hence interacts well with electron donors or Lewis bases generating coordination complexes.[40] SO_3 forms a variety of adducts with amines and the basicity of the amine affects the stability and the reactivity of the SO_3 complex. The more basic the amine is, the more stable and less reactive the complex is. Anyway, even the weakest complex is a much milder reagent than free Sulphur trioxide. The complex with pyridine is quite stable, though reactive enough to sulfate the hydroxyl groups of carbohydrates without degrading the sugar chains.

Sulfur trioxide-pyridine complex **35** can be prepared in situ by reaction of chlorosulfonic acid with pyridine before the addition of the carbohydrate.

$$ClSO_3H + 2\ py \longrightarrow SO_3{\cdot}py + pyHCl$$
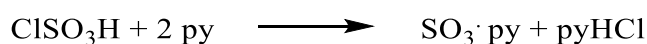
Preliminary experiments on this step were performed at room temperature in a flask equipped with a classical dropping funnel under nitrogen flow and connected to two NaOH scrubbers. We observed a strong increase in temperature (even 10°C after each drop) and the formation of puffs with formation of solid material which will block the nitrogen flow to the scrubbers. The puffs can be due to instantaneous development of gaseous compounds (e.g. HCl or pyHCl) which determines a change in the volume of the head space.

To avoid this problem chlorosulfonic acid was added under the surface of pyridine in order to avoid the formation of the eventual hot-spots on the surface and hence to prevent releasing of either HCl or pyHCl in the head space. The addition was carried out through a peristaltic pump and a Teflon tube placed under the level of pyridine (Figure 3.18). The dosage rate and the temperature of the vessel jacket were set to have an internal temperature between 50 and 60 °C. It is important to keep the temperature above 50°C to avoid precipitation of a high amount of py·SO3 complex, which would complicate the stirring of the reaction mixture.
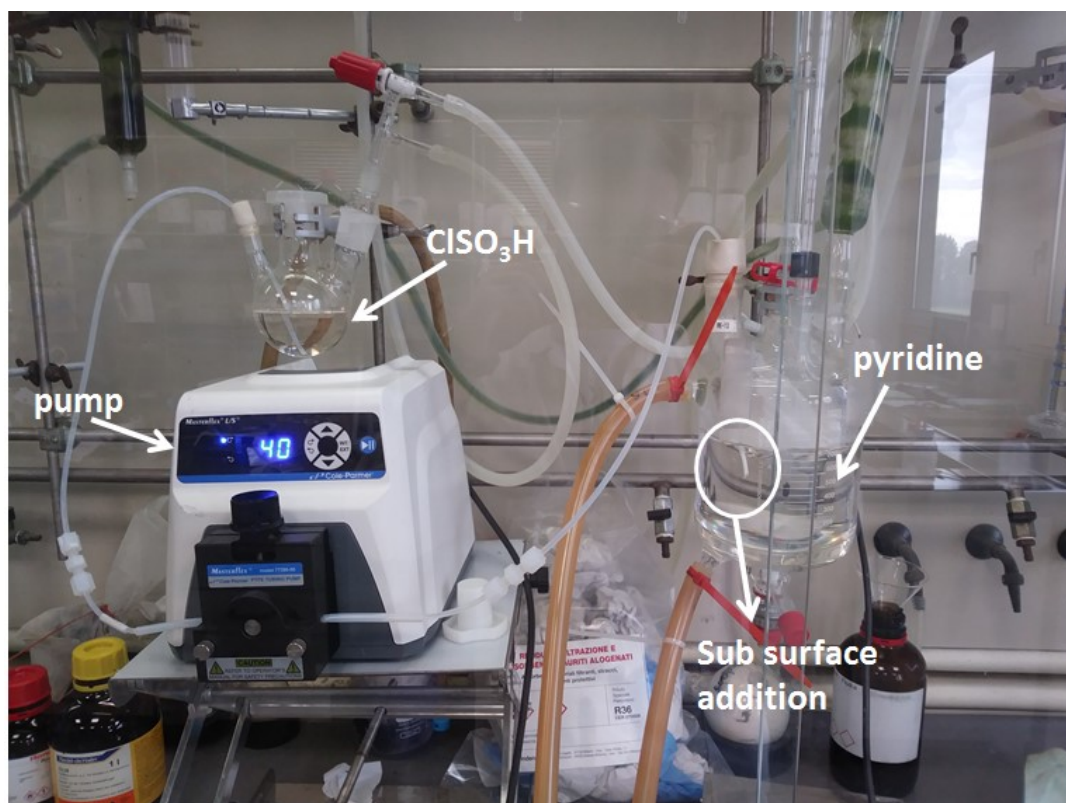
Figure 3.18. Experimental apparatus for the formation of sulfur trioxide-pyridine complex.

This system allows to add chlrorosulfonic acid to pyridine, while keeping the temperature under control and avoiding the formation of puffs and white solid on the walls of the vessel.

At the end of the addition, the mixture is heated at 60°C for 30 min to complete the formation of the adduct. At this stage the xylan is introduced, sulfation occurs and the product is isolated by

precipitation in methanol. The isolated product has pyridinium groups as counter-ions of the sulfates, while the target compound is commercialized as sodium salt. To exchange the cation, the solid was dissolved in water and treated with aqueous NaOH until basic pH. After neutralization of the aqueous solution the crude sodium salt of the sulfated polysaccharide is isolated by precipitation in methanol.

^{1}H-NMR is a crucial analysis to control the outcome of this step because it allows to evaluate roughly the degree of sulfation. In Figure 3.19 a comparison of the proton spectra of the glucuronoxylan before sulfation, a partially sulfated xylan (obtained with sub-stoichiometric amount of ClSO$_3$H) and a product with the desired sulfate content is shown. The signals of the protons of a sulfated xylose (red) are shifted to lower field than the protons of a xylan (black). In the case of an under-sulfated product (blue) an intermediate situation is observed characterized by signals of both sulfated and not sulfated xylose units. This analysis does not provide a quantitative evaluation, but it can be a quick in-process control (IPC) to know if the sulfation degree is close to the target. A more accurate determination of the sulfur content can be achieved by ICP-MS or the other methods discussed in paragraph 2.3.
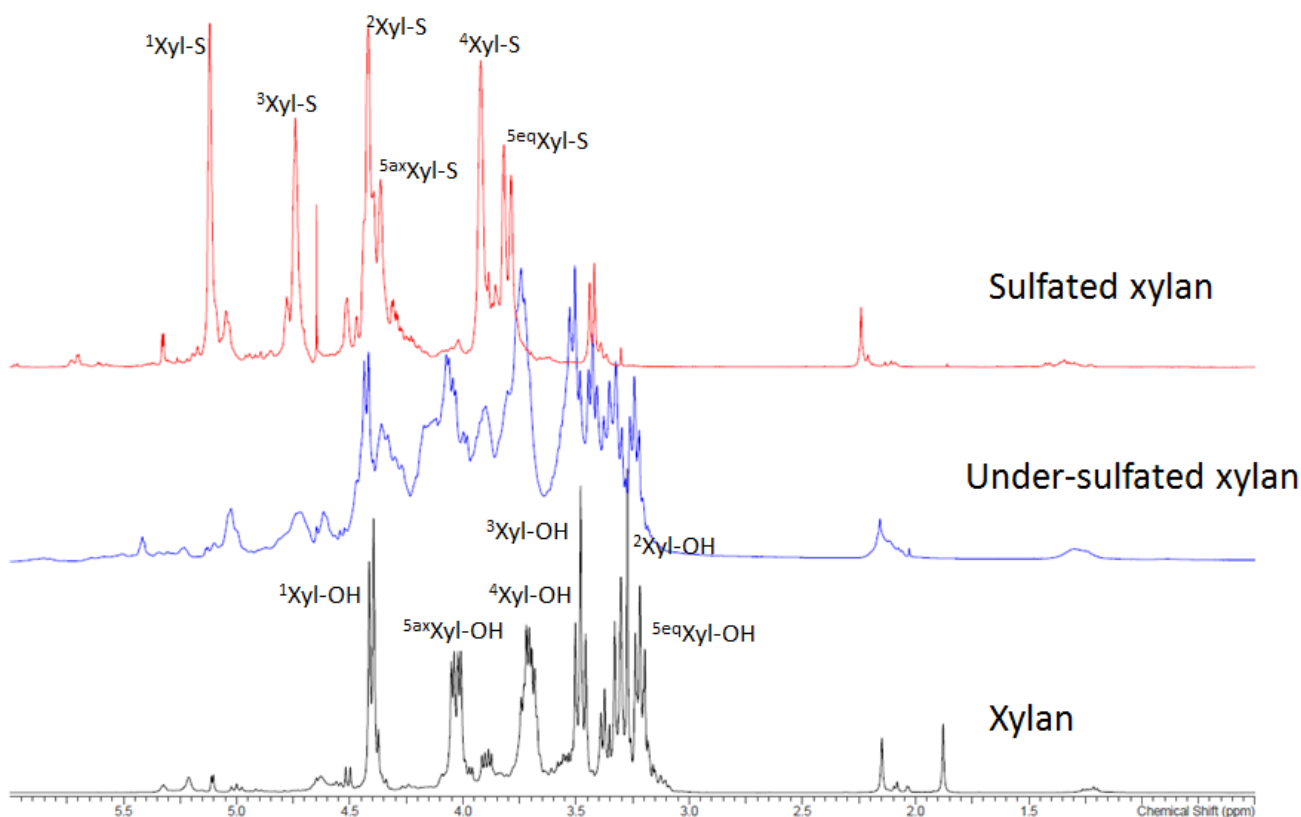

Figure 3.19. Comparison of ^{1}H-NMR spectra of xylan with different degree of sulfation.

Furthermore, the ^{1}H-NMR spectrum of crude sulfated polysaccharide is useful to evaluate other aspects highlighted in Figure 3.20. In the region between 7.5 and 9.5 ppm (red area) the signals of pyrdinium moieties and pyridine are shown. Pyridinium groups are attached to the anomeric position of xylose units at the reducing end of the chain, structure **3**, and are a typical feature of the Originator SP, as discussed in paragraph 2.2. These species are obviously generated during the sulfation step. In order to have the correct amount of these species in the SP, the equivalents of chlorosulfonic acid and the reaction temperature are important. Indeed, a decrease of the equivalents of ClSO$_3$H used and especially the reaction temperature determines a reduction of the number of pyridinium groups **3**. At the same time, these parameters affect the degree of sulfation and hence it is necessary to find conditions which avoid under-sulfation and at the same time allow to obtain the right content of oligomers **3**. We found that the amount of chlorosulfonic acid must be maintained between 1.2 and 2.0 mL per gram of xylan, which correspond to about 1.2-2.0 equivalents per hydroxyl group according to a rough calculation. The reaction temperature is the most significant parameter and needs to be set between 60 and 80°C, preferably at 70°C. In the same spectral region another group of three signals related to pyridine, which is not bound to the product chain, are visible.
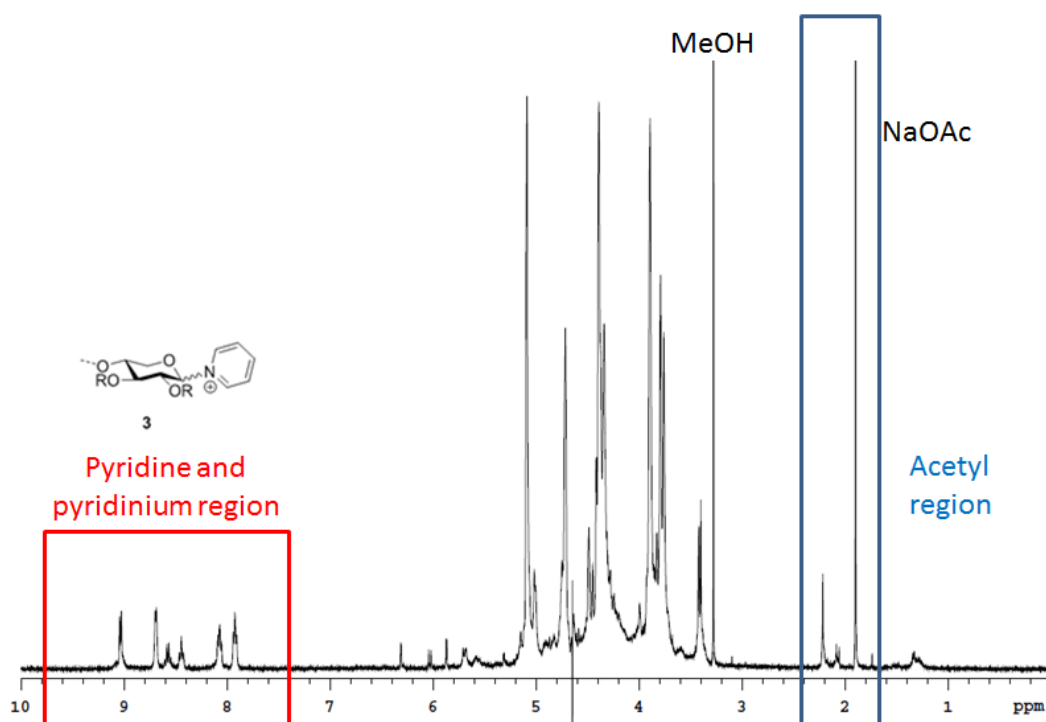

Figure 3.20. ^{1}H-NMR spectrum of crude sulfated polysaccharide.

In the blue region at about 2 ppm the group of signals related to acetyl groups appear. It is important to monitor them because the formation of the sodium salt of SP after the sulfation step

can be critical for acetates. Indeed, as explained before, aqueous NaOH is used to exchange the sulfate counterion from pyridinium to sodium and in basic conditions deacetylation can occur. To avoid that, pH must be maintained between 10.5 and 12.5 and the temperature below 25°C. Moreover, in the same region there is an intense signal due to the presence of sodium acetate which is generated in the neutralization of NaOH solution with acetic acid.

The crude sulfated product was also analyzed by GPC-RI and the typical chromatogram is reported in Figure 3.21. The usual values of our crude products (Mn = 3.5-5.0 KDa Mw = 5.8-9.0 KDa, MP = 5.1-7.1 KDa) are close to the range obtained, with this analytical method, for the originator's sulfated polysaccharide (Mn = 3.8-5.7 KDa, Mw = 7.8-8.4 KDa, MP = 5.1-5.9 KDa). However, the results are not very accurate, because the product is contaminated by inorganic salts which modify the shape of the peak and consequently the values obtained by integration. The main inorganic impurity is sulfate ions, which show up in the tail of the main peak.

Mn	MP	Mw	PD
4681	7081	9015	1,9
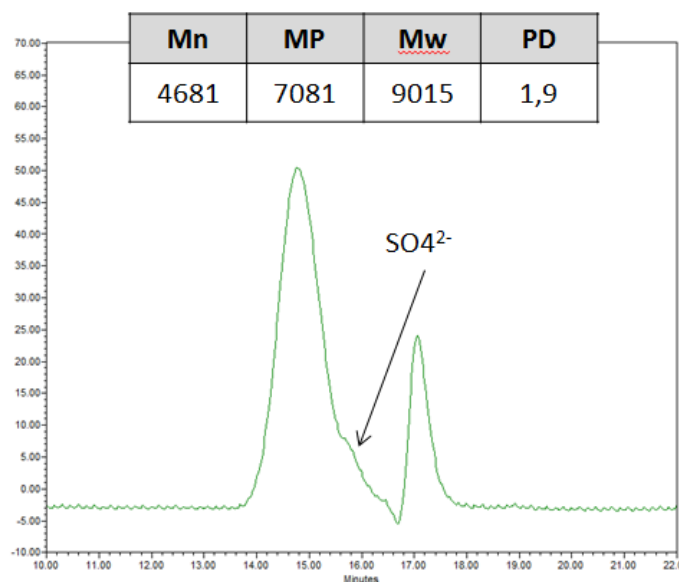
Figure 3.21. Typical GPC-RI chromatogram of crude sulfated polysaccharide.

In conclusion, after sulfation the skeleton of the SP is completed, but the product is contaminated by residual solvents (e.g. pyridine, methanol) and inorganic salts (e.g. NaOAc, Na_2SO_4) which must be removed. Moreover, the product is a brown solid and hence it will require a bleaching step to obtain the white color proper of the Originator SP.

3.9 Product purification: Dialysis and TFF

To remove residual salts and solvents, we evaluated different methods and one of them is dialysis, which has been extensively used for the purification of this class of products.[34] Dialysis is a separation process based on the diffusion of molecules through a semipermeable membrane. The

membranes have pores, which let the smallest molecules pass through and retain the biggest species. We used the so-called Dialysis cassettes (Figure 3.22), which are characterized by hermetically sealed chambers with membranes of low-binding regenerated cellulose. The cassettes have different molecular weight cutoff (MWCO), which is the maximum molecular weight of the molecules which diffuse through the membrane. It is an estimation because the passage is not determined by the molecular weight, but by the size of the molecules and hence the chemical and conformational properties play a key role. It can happen that a molecule with a molecular weight higher than MWCO passes through the membrane. Considering the molecular weight of our product (4000-9000 Da) and of the species to remove (NaOAc, Na_2SO_4, py), we chose the cassettes with MWCO of 2 KDa.


Figure 3.22. Dialysis cassettes.

First of all, the cassette is immersed in water for about 10' to be hydrated and then is filled with a solution of the crude product through a hole placed in a corner. The air is removed from the inner part by a syringe to increase the surface of the liquid and the cassette is soaked in water, which is changed after a certain time (4-20 h). Conductivity can be used to monitor the dialysis course. Indeed, the conductivity is the ability of an electrolyte to conduct electricity and depends on the concentration of the electrolytes in solution. Thus, the movement of charged species, like the inorganic salts, from the inner of the cassettes to the external water should raise the conductivity of the external water. Thanks to these measures we could set the dialysis protocol and we charged 200 mg of crude product dissolved in 20 mL of water. Then, we immerged the cassettes in 1.8 L of water and we changed the water after 6 h, 22 h and 30 h measuring the conductivity. Typical conductivity values are reported in Table 3.7 where we can see that in the first and second external waters the conductivity is significantly higher than the value of distilled water. On the other hand, in the third external water the conductivity increase is minimal, which is an indication that almost all the inorganic salts present in the crude have been removed.

Sample	t (h)	Conductivity (µS/cm)
Distilled water before use	0	6.20
First external water	6	18.51
Second external water	22	21.24
Third external water	30	9.04

Table 3.7. Conductivity measure of dialysis waters.

Therefore, after 30 h we can consider the dialysis complete and we can freeze dry the content of the cassette to obtain the pure sulfated polysaccharide (recovery = 60-70%). NMR and GPC-RI confirm that all the inorganic salts and residual solvent are efficiently removed with dialysis.

The main disadvantage of this purification system is that it is not scalable and transferable to an industrial plant. However, there is a similar, but industrializable method called tangential flow filtration (TFF), which is sketched in Figure 3.23. The sample is dissolved in water and it is forced by a peristaltic pump to flow through a capsule, which has a membrane similar to the one used in dialysis. Indeed, the molecules with lower size pass through the membrane pores and go into the so-called filtrate or permeate, while the biggest species cannot cross the membrane and remain in the so-called retentate. The latter goes back to the feeding solution which is continuously pumped by the peristaltic pump in TFF capsule.
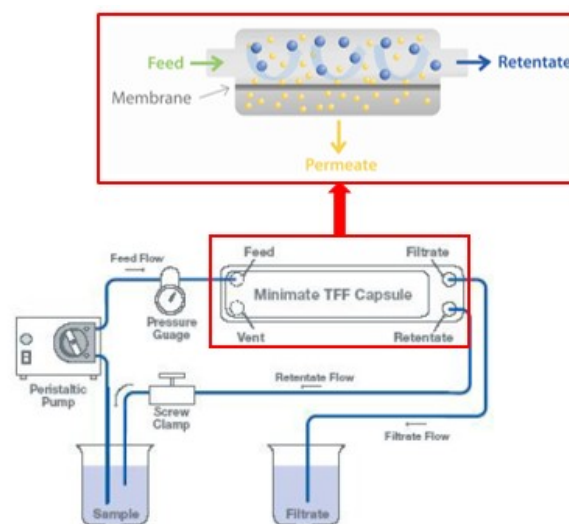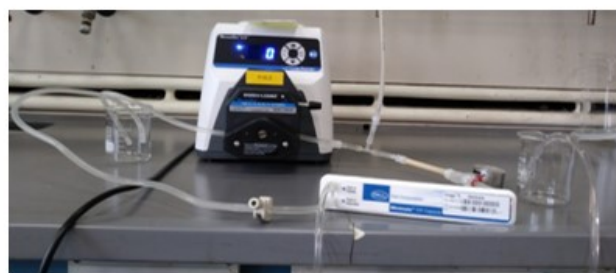
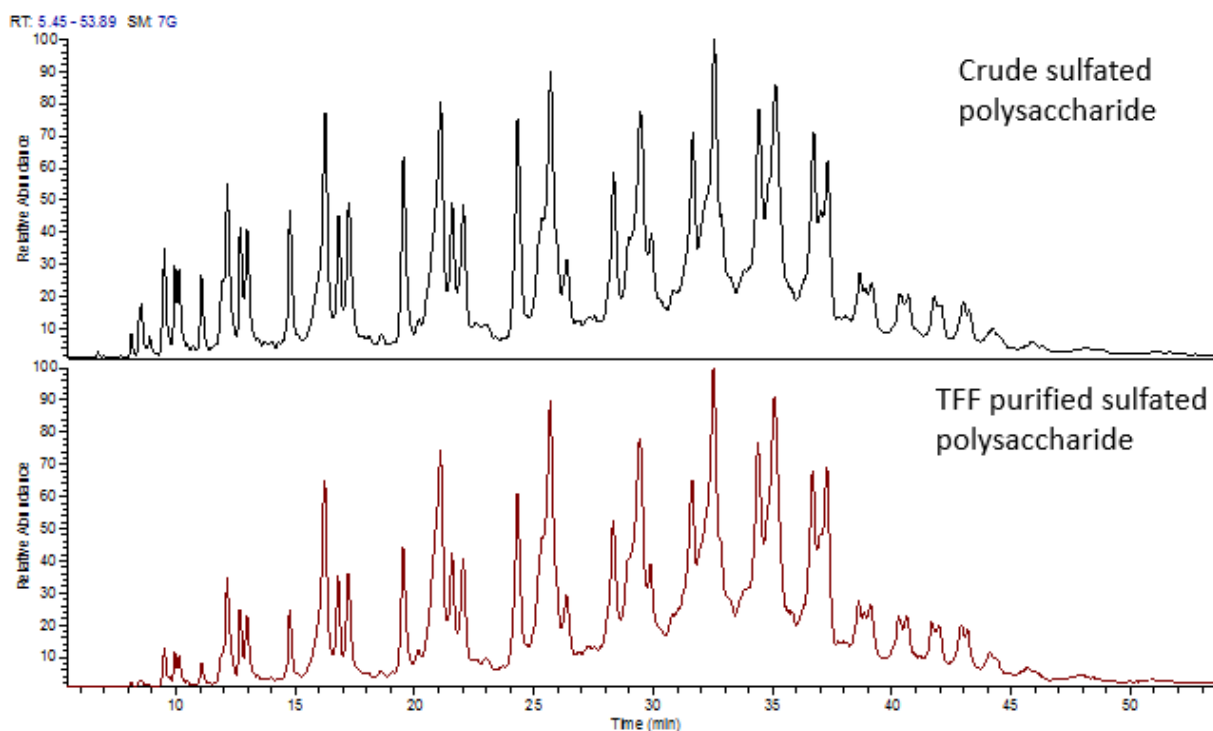**Figure 3.23. Tangential flow filtration (TFF) system.** (Figure modified from Reference 41).

The tangential flow of the solution prevents the formation of fouling on the surface of the membrane and hence avoids a decrease in the filtration performance and in turn an increase in the pressure of the system over time.

TFF can be exploited to concentrate or to diafiltrate a solution. The process is the same but when you want to diafiltrate you add fresh water or buffer to the sample solution in conjunction with concentration. Diafiltration can be continuous or discontinuous. In the first case fresh water is added at the same rate of the filtrate flux so that the volume of the retentate does not change. In the second case the solution is first concentrated and then diluted with fresh water to its original volume. This sequence is repeated several times. The continuous mode of operation is more efficient, but in both the methods after addition of seven volumes of diafiltration the ionic strength decreases more than 99%[41] (One diafiltration volume equals the initial volume in which the molecule of interest is dissolved).

Like in dialysis, the membranes can have different molecular weight cut off (MWCO) and we selected devices with MWCO of 1 KDa. It is important to notice that the material of TFF membranes (modified polyethersulfone) is different from dialysis membranes (regenerated cellulose) and hence the behavior of the compound in the two systems can be different due to different chemical interactions.

Once chosen the MWCO of the capsule, there are still two main variables to optimize: transmembrane pressure (TMP) and the flow rate. The transmembrane pressure is the difference between the pressure on the two sides of the membrane. To measure TMP in our system we have set a pressure gauge on the inlet side of the capsules and we approximate the pressure at the outlet side equal to the atmospheric pressure. The transmembrane pressure can be controlled by the screw clamp placed on the retentate tube. In our experiments we have used TMP equal to 20-30 psi and a flow rate of 0.1 L/h, which once normalized was equal to $J = 0.3$ L/ m^2 min.

At the end of TFF the sample solution was freeze dried finding that the salts and the residual solvents were removed efficiently. Indeed, ^{1}H-NMR spectra did not show signals of pyridine, methanol and sodium acetate and the GPC-RI chromatograms did not show the sulfate peak in the tail of the SP peak. The recovery obtained by TFF (50-60%) was slightly lower than the one obtained with dialysis (60-70%), yet the purification was not completely selective because a portion of the SP managed to go through the membrane. The comparison of LC-MS chromatograms shows apparent differences between the crude SP (black) and the SP (brown) purified by TFF (Figure 3.24).

Figure 3.24. Comparison of LC-MS profiles of crude and TFF purified product.

The peaks at low retention time have lower relative intensity in the product after tangential flow filtration. These peaks belong to oligomers with low degree of polymerization which probably went through the membrane during TFF. Indeed, their molecular weights are really close to molecular weight cutoff of the membrane. To confirm this data, we also collected and freeze dried the permeate. GPC analysis revealed a molecular weight significantly lower than the crude SP (Mp crude = 7 KDa, Mp TFF permeate = 3 KDa). Therefore, we can conclude that the purification carried out with TFF affects the composition of the product and causes the loss of species with low degree of polymerization. These oligomers are a key feature of the Originator SP and are present with intense peaks in its LC-MS chromatogram.

In conclusion, dialysis is not scalable on an industrial scale and TFF changes the composition of the product making it less similar to originator's one. We needed to find an alternative method to purify the product and we evaluated a method simply based on precipitations.

3.10 Product purification by precipitation

To remove the residual pyridine and the salts (Na_2SO_4 and NaOAc), reprecipitation of the crude product might be a simple and effective strategy. However, there is an issue linked to sodium sulfate which is present in high amount and is not removed efficiently by simple reprecipitation. Indeed, its solubility in methanol and water (s H_2O (20°C) = 139 g/L; s MeOH (20°C) = 24.6 g/L) is

176

not as high as that of sodium acetate (s H_2O (20°C) = 1230 g/L; s MeOH (20°C) = 160 g/L). Therefore, we developed another method for sulfate removal which is based on the addition of barium chloride to a water solution of the crude product.[42] Barium chloride reacts with sodium sulfate to afford sodium chloride and barium sulfate which is almost insoluble in water (s H_2O (20°C) = 2.5 x 10^{-3} g/L) and hence precipitates.

$$Na_2SO_4 + BaCl_2 \longrightarrow 2\ NaCl + BaSO_4$$
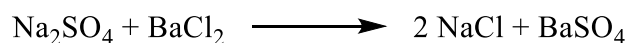
Barium chloride was added to the crude sulfated product in water until the formation of precipitate was complete in order to remove all the inorganic sulfates. Then, the suspension was filtered and the filtrate was concentrated and finally methanol was added to isolate the pure SP, which precipitates. We ran ICP-MS analysis on the obtained product to determine the sulfur, sodium and barium content and we found an unexpected result (Table 3.8). The pure SP obtained after $BaCl_2$ treatment has low values of sulfur and sodium and a high content of barium. These results could be interpreted as if desulfation of the polysaccharide chain had occurred, but ^{1}H-NMR spectrum shows the typical profile of SP with the correct degree of sulfation. A more accurate evaluation of the results allowed to appreciate that the product was not desulfated by barium addition. ICP-MS analysis can be perturbed by the presence of high amounts of barium because, according to ICP-MS analytical method, the sample is first mineralized and then filtered before the introduction in the ICP-MS equipment. However, during the mineralization step the sulfate groups bound to the chain are cleaved and interact with barium present in the product to generate barium sulfate, which precipitates and is hence filtered off before the analysis. The result is that a portion of sulfates are removed during sample preparation and thus the content of sulfur and barium is underestimated. On the other hand, the value of sodium is not compromised but it is nevertheless very low compared to crude product and to the Originator SP. Sodium is probably replaced by barium as counterion of sulfate groups, as suggested by its high value.

Sample	S % (w/w)	Na % (w/w)	Ba (ppm)
Crude SP	18.2	13.9	158
Pure SP	7.8	1.0	13441
Originator SP*	15.5	12.0	2-4

Table 3.8. Sulfur, sodium and barium content of crude, pure (after BaCl₂ treatment) and Originator SP.
* the values of the Originator SP are reported as average of ten batches.

In conclusion, ICP-MS analysis is not reliable when the product is contaminated by high amounts of barium. Anyway, ^{1}H-NMR suggests that the addition of barium chloride to crude product does

not alter the structure of the polysaccharide, but changes instead the counterion of sulfates from sodium to barium. Therefore, a method is needed to regenerate the sodium salt and to decrease barium content to ppm units (Table 3.9). First of all, we tried to switch the cation by simply treating a water solution of the sample with NaOH (Exp. 1) or adding dropwise a SP water solution to NaOH in methanol (Exp. 2). However, in both the experiments the content of Barium was still above 6000 ppm. Another strategy was to precipitate the excess of barium by addition of sodium sulfate (Exp. 3) and sodium carbonate (Exp. 4), which could exchange the cation of the sulfated polysaccharide chain. Sodium carbonate was more effective than sodium sulfate but Ba content did not decrease below 900 ppm. We found only a method able to decrease the content of Barium at ppm units and that was based on the use of a cation exchange resin, specifically Amberlyst 15 Dry (Exp. 5). This resin has styrene and divinylbenzene matrix functionalized with sulfonic acid groups and is usually commercialized in the acidic form, but can be conditioned with NaOH or NaCl solutions to exchange H^+ with Na^+. When the SP aqueous solution is passed through the resin, it exchanges barium for sodium. This method allows to decrease barium content to about 4 ppm and to obtain sulfur and sodium levels very close to the values of the Originator SP.

Exp	Method of Ba removal	S % (w/w)	Na % (w/w)	Ba (ppm)
0	Starting SP	7.8	1.0	13441
1	NaOH in water	9.7	5.9	8980
2	NaOH in MeOH	11.9	8.4	6192
3	Addition of Na_2SO_4	19.3	19.9	7982
4	Addition of Na_2CO_3	11.1	22.0	966
5	Cation exchange resin	17.1	10.8	4

Table 3.9. Experiment of Sodium salt regeneration and barium removal. The content of sulfur, sodium and barium was determined by ICP-MS.

In Exp. 5 we used 25 volumes of wet resin per gram of crude product and hence a large excess of resin and cation exchange sites. In the future, we will optimize the method to decrease the amount of resin to the lowest possible limit. Moreover, we will need to define a protocol for the regeneration of the resin in order to be able to re-use it. The big affinity of barium with sulfonic sites complicates the regeneration of the resin and high concentration of NaOH or NaCl in the regenerating solution will probably be necessary.

The product eluted from the resin is free of sulfates and barium and is concentrated to low volumes (2-3 V vs dry residue). At this stage some white crystals usually crash out, probably

sodium chloride generated by the addition of BaCl$_2$. The solid was filtered off and the concentrated aqueous solution was precipitated in methanol to isolate the pure SP.

This strategy based on three main steps (treatment with BaCl$_2$, cation exchange resin, final precipitation) is efficient in the removal of salts and residual solvents. Indeed, ^{1}H-NMR analysis does not show the signals of pyridine and sodium acetate and GPC-RI does not present the sulfate peak in the tail of the SP peak. Moreover, GPC-RI demonstrates that the purification steps does not alter the molecular weight distribution, as shown in Table 3.10.

Sample	Mn	Mw	PD
Crude SP	5074	6636	1.3
Pure SP	5150	6776	1.3

Table 3.10. Comparison of GPC-RI results before and after purification.

These results are also supported by LC-MS chromatograms reported in Figure 3.25. Indeed, the profiles before and after purification are very similar and there are no significant differences between the relative intensities of the signals. The analysis suggests that this method, in contrast to tangential flow filtration (TFF), does not cause the loss of the low molecular weight fragments which are also present in the Originator SP.
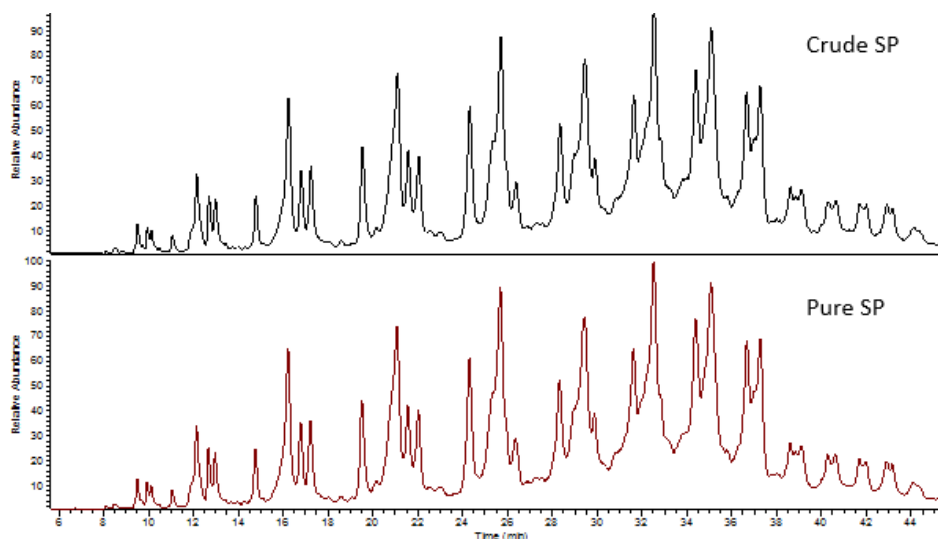
Figure 3.25. Comparison of LC-MS chromatograms before and after purification.

In conclusion, we developed a method of purification of the sulfated polysaccharide characterized by three main steps: removal of sodium sulfate by precipitation with BaCl$_2$, elution through a cation exchange resin to get rid of the excess of barium and finally, precipitation in methanol to isolate the product. This protocol allows to remove efficiently inorganic salts and residual pyridine.

Moreover, it is scalable and does not have a negative impact on the quality and the key structural features (e.g. the low molecular weight oligomers) of the SP.

3.11 Bleaching of sulfated polysaccharide

All the products obtained with the process described until now (extraction of glucuronoxylan from beech wood, sulfation and purification) have a brownish color, while the originator's sulfated polysaccharide is a white solid. Therefore, to achieve similarity also on this aspect, it is necessary to bleach our SP. Moreover, the white color is crucial to obtain good results in GPC-MALLS analysis. Indeed, as explained in paragraph 2.1, MALLS (Multi Angle Laser Light Scattering) detector determines the molecular weight according to the scattering of the light emitted by a laser. The detector does not provide reliable results with brownish samples. This issue is linked to the presence of fluorescent impurities which can emit radiations at the same wavelengths of the MALLS detector laser. Thus, the white color of the product is crucial to obtain the correct molecular weight values in GPC-MALLS, one of the key analytical techniques.

First of all, to whiten the sulfated polysaccharide we tried adding charcoal to the sample solution in water until 10% w/w, but no change of color was observed both at room temperature and under reflux. Therefore, we explored the use of oxidants such as hydrogen peroxide, which requires at least 50°C to be effective at low concentrations (≤ 5%). Bleaching can also be achieved by 3-5% sodium hypochlorite or chlorine dioxide generated in situ by addition of acetic acid to a solution of sodium chlorite, as explained in paragraph 3.4. To avoid safety issues related to its explosive properties, chlorine dioxide should be used below 1% w/V in solution. These three different oxidants were evaluated in their bleaching abilities starting from the same sulfated polysaccharide and comparing the color of the solids obtained after bleaching (Figure 3.26). The not bleached product has a brownish color which turns into yellowish after bleaching with hydrogen peroxide and chlorine dioxide. On the other hand, sodium hypochlorite treatment yields an off-white solid and thus seems the most promising method.

Figure 3.26. Comparison of samples obtained with different bleaching agents.

Furthermore, these samples were also analyzed to understand the impact of bleaching steps on the structure of the sulfated polysaccharide. ^{1}H-NMR spectra do not show evident changes, while GPC-RI shows a slight decrease of the molecular weight (Table 3.11). All the whitening methods decrease Mn e Mw of the SP and NaClO increases also the polydispersity.

Sample	Mn	Mw	PD
Unbleached SP	4111	7527	1.83
SP bleached with H_2O_2	3882	7109	1.83
SP bleached with ClO_2	3798	7112	1.87
SP bleached with NaClO	3717	7247	1.95

Table 3.11. GPC-RI results of bleaching experiments.

Finally, a very accurate comparison of the whitening methods was achieved by LC-MS (Figure 3.27). The analysis suggests that hydrogen peroxide and sodium hypochlorite do not affect LC-MS profile, which is exactly the same of the unbleached sample. On the other hand, after treatment with chlorine dioxide it is evident a new series of peaks not present in the starting SP and marked with "?'" symbol in Figure 3.27. We were not able to determine the structure of these oligomers, but LC-MS shows that they are not present in the Originator SP.
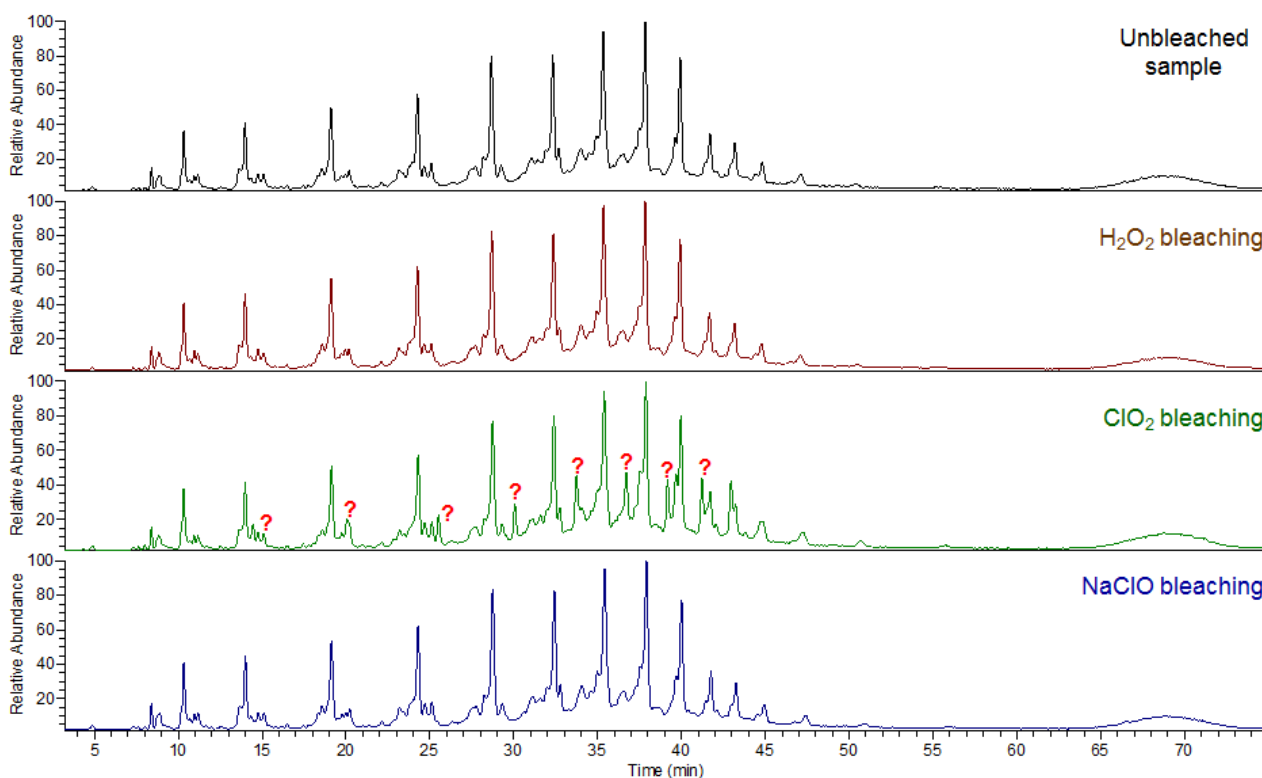

Figure 3.27. LC-MS profiles of bleaching experiments.

In conclusion, we studied the use of three oxidants as whitening agents and among them sodium hypochlorite seems the most promising one because it affords white or off-white products even when used at low concentration. Moreover, it does not affect the structure of the product introducing new chemical species absent in the Originator SP.

3.12 Process conclusion and future development

The main steps of the process for the preparation of the sulfated polysaccharide were studied and defined, as summarized in Scheme 3.7. First of all, we investigated the extraction of the glucuronoxylan from beech wood and we evaluated three alternative methods. Among them autohydrolysis seems the most promising one because it does not require any prior delignification and the use of hazardous reagents. According to this method, the glucuronoxylan is simply extracted from beech wood by water at high temperature (120-150°C). Moreover, the xylan is depolymerized during the extraction and thus a further depolymerization step is not necessary. After autohydrolysis the xylan is selectively deacetylated with a slightly basic aqueous solution. We optimized the autohydrolysis and deacetylation conditions to obtain the glucuronoxylan with the desired molecular weight, acetyl groups number and all the key structural features. This part of the process was scaled-up in pilot plant to prepare 1 Kg of xylan. In the future, the process will be further optimized in order to reduce the volumes of extraction and precipitation and consequently to increase the volumetric productivity.

Then, the glucuronoxylan was sulfated by a complex of sulfur trioxide with pyridine, which is generated in situ by addition of chlorosulfonic acid to pyridine. This step is highly exothermic and generates vapor puffs and hence we modified the dosage mode and the rate of $ClSO_3H$ addition developing an efficient and safe method for multi grams preparation. However, in future an accurate calorimetric study will be carried out to guarantee the safety of the process in an industrial plant. After sulfation, the product is first isolated as pyridinium salt and then treated with NaOH to generate the corresponding crude sodium salt. The amount of organic solvent for the precipitation of the crude sulfated polysaccharide is very high and experiments for its reduction will be carried out. Finally, the product needs to be bleached and purified. We developed an efficient and scalable method based on addition of $BaCl_2$, cation exchange resin and final precipitation in methanol.
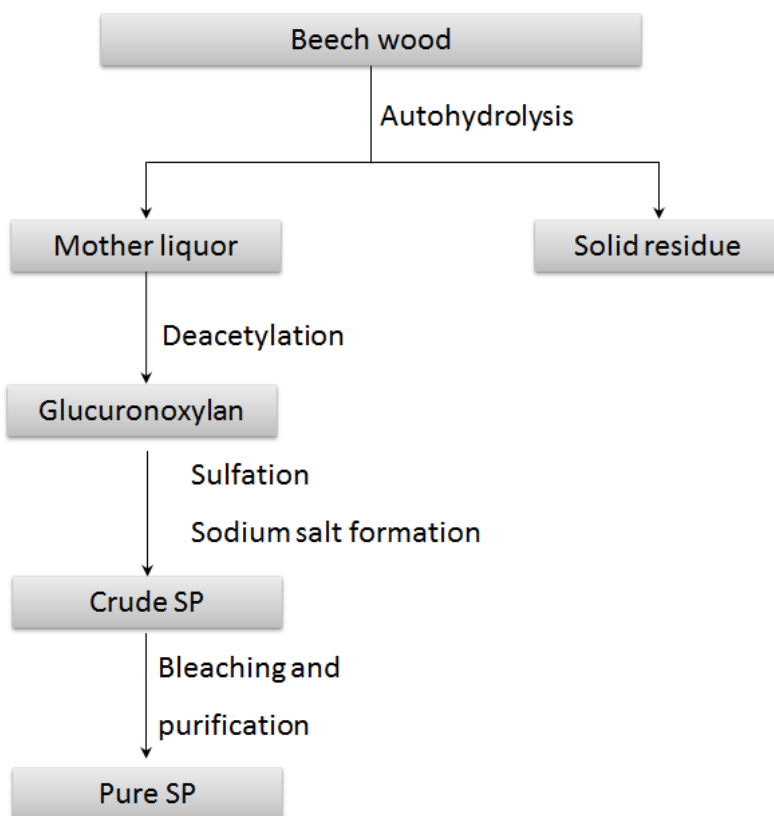
Scheme 3.7. Schematic representation of the process defined for the preparation of the sulfated polysaccharide.

In conclusion, the process needs to be optimized before industrialization to increase the volumetric productivity, but the key steps from beech wood to sulfated polysaccharide have been defined. The product obtained by this process is similar to Originator's version. A complete comparison of the developed SP and the Originator SP will be presented in the next chapter.

Experimental section

Some experimental procedures are not reported for confidentiality reasons.

Beech wood sawdust was purchased from two different sawmills, but the names and the countries of the producers are not reported for confidentiality reasons. The chemical reagents ($NaClO_2$, $Na_2S_2O_3$, 38-40% AcOOH, 30% H_2O_2, $ClSO_3H$ and $BaCl_2$) were purchased from Sigma Aldrich – Merck, while the solvents were technical grade solvents.

Nuclear magnetic resonance spectra were obtained on a Agilent MR400 DD2 VNMRJ™ Software Rev. 4.2. All ^{1}H-NMR experiments were recorded at 400 MHz and at a temperature of 30 °C. For NMR analysis of glucuronoxylan about 10 mg of sample were dissolved in 0.75 mL of D_2O, while for the sulfated products about 75 mg of sample were dissolved in 0.75 mL of D_2O.

GPC-RI method was discussed in chapter 2, but the experimental details cannot be described for confidentiality reasons. The technique has an intrinsic variability between different analytical sessions so that the Originator SP was always used as a reference. The absolute values of a sample can be slightly different, but it is possible to evaluate in a relative way the differences in the molecular weight introduced by a step of the process.

Chlorine dioxide delignification of beech wood

In a 2 L flask 50 g of beech wood sawdust (water content = 8%, 4.0 g) were suspended in 500 mL (10 V) of water under mechanical stirring and nitrogen flow. The flask was connected to two scrubbers containing $Na_2S_2O_3$. The mixture was heated at 55-60°C and AcOH was added (2.5 mL, 2.6 g, 44 mmol). A solution of 13.3 g of 80% $NaClO_2$ (10.6 g, 118 mmol) in 17.7 mL of water was added in 5 min. and the mixture was stirred at 55-60 °C for 1 h. AcOH (2.5 mL) and a solution of 13.3 g of 80% $NaClO_2$ in 17.7 mL of water were added a second time and the mixture was stirred again for 1 h at 55-60°C. The addition of AcOH (2.5 mL) and of the solution of 80% $NaClO_2$ (13.3 g) in water (17.7 ml) was repeated for the third time and the mixture was stirred for an additional hour at 55-60°C. The suspension was cooled to room temperature and filtered. The solid was washed with 10 x 500 mL of water, 1 L of 96% EtOH and 1 L of MeOH. Finally, the wet solid (101.3 g) was dried under vacuum at 50°C for 20 h to yield 41.2 g of white beech wood sawdust.

Loss of weight vs starting sawdust = 19.6 %

Loss of weight vs dry sawdust = 10.5 %

DMSO extraction of xylan

In a 250 mL flask 10.0 g of chlorine dioxide delignified sawdust were suspended in 100 mL (10V) of DMSO under magnetic stirring and a nitrogen flow. The mixture was stirred at 60°C for 24 h turning to a brownish color. The suspension was cooled at room temperature and filtered. The solid was washed with 2 x 10 mL of DMSO. 80 mL of filtrate were collected. 320 mL of absolute EtOH and 238 μL of 37% HCl were added and the color of the solution turned from brown to pale yellow. The mixture was cooled at 0-5 °C for 4 h and a precipitate was observed. The suspension was filtered and the solid was washed with 2 x 10 mL of EtOH. The solid was dried under vacuum at 50°C for 40 h to yield a brownish solid (25 mg, yield (w/w) = 0.25 %).

GPC-RI: the high content of inorganic salts does not allows the detection of xylan peak

^{1}H-NMR (D$_2$O):
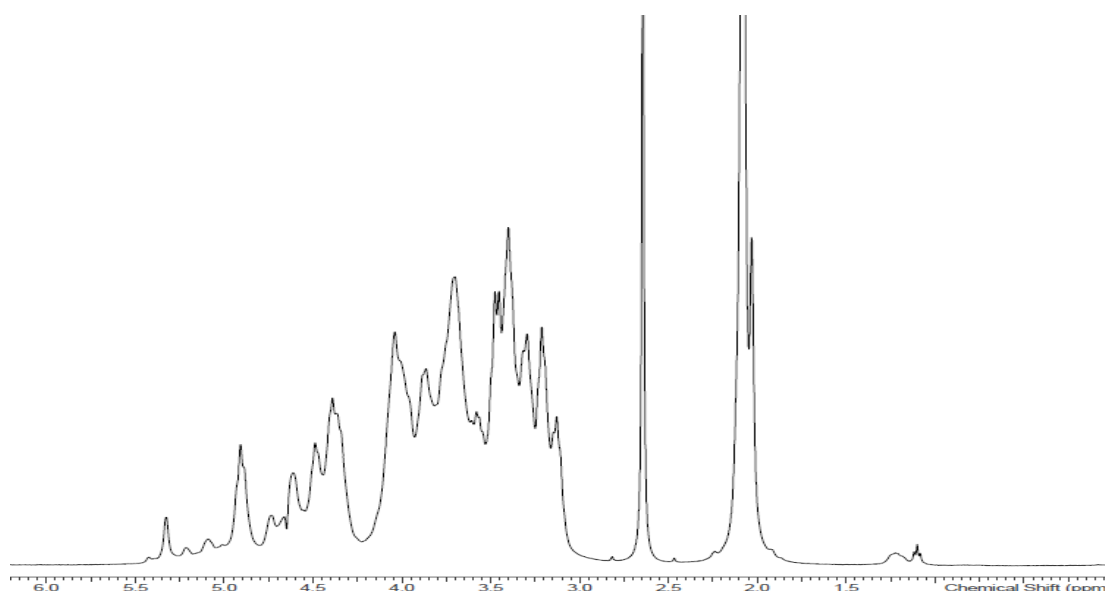
Peracetic acid delignification of beech wood

In a 3L flask 242 g of beech wood sawdust (water content = 7.9%, 19.12 g) were suspended in 1.37 L (5.6 V) of water under mechanical stirring and a nitrogen flow. The mixture was heated at 80-85 °C and 562 g of 38-40 % w/w AcOOH solution (219 g of AcOOH, 2.88 mol) were added in 3 h. At the end of the addition the mixture was cooled at room temperature. The suspension was filtered and the solid was washed with 6 x 4 L of water, 3 x 1.6 L of 96% EtOH and 3 x 1.6 L of MeOH. The solid was dried under vacuum at 50°C for 20 h to yield 128 g of white beech wood sawdust.

Loss of weight vs starting sawdust = 47.0 %

Loss of weight vs dry sawdust = 42.4 %

Ammonia extraction of xylan

120 g of delignified beech wood sawdust were suspended in 2.4 L (20 V) of 1% of NH_3 solution in water under mechanical stirring and a nitrogen flow. The mixture was stirred at room temperature for 4 h and then filtered. The solid was washed twice with 300 mL of water and resuspended in 2.4 L of 1% of NH_3 solution under mechanical stirring and nitrogen flow. The suspension was stirred at room temperature for 4 h, filtered and the solid was washed with 2 x 300 mL of water. The two filtrates (pH = 10.5-11) were neutralized with acetic acid, merged and concentrated at reduced pressure to 240 mL (2 V). The brownish solution was added drop-wise to 840 mL (7V) of 96% EtOH with 8.4 g (1% w/V) of AcOH. The suspension was stirred at 0-5 °C for 20 h and filtered. The solid was washed with 3 x 240 mL of cold 96% EtOH and dried under vacuum at 50°C for 20h.
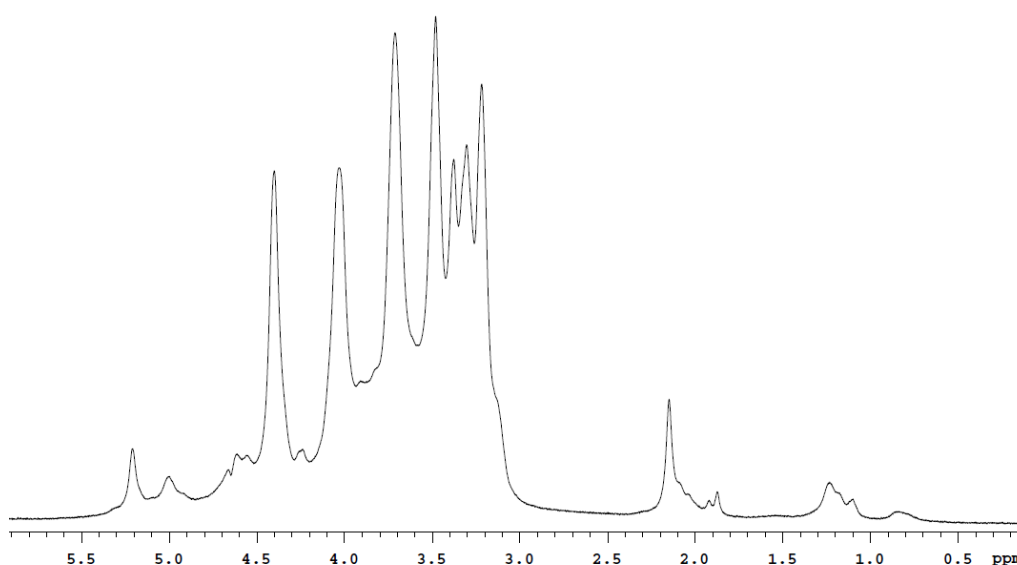
Xylan from ClO_2 delignfied sawdust: 3.26 g (yield vs delignified sawdust = 2.71%; yield vs starting sawdust = 2.25%).

Xylan from AcOOH delignfied sawdust: 9.04 g (yield vs delignified sawdust = 7.53%; yield vs starting sawdust = 4.0%).

GPC-RI:

| Delignification Method	Mn (Da)	Mw (Da)	MP (Da)	PD
ClO_2	2956	12426	11292	4.2
AcOOH	9583	28470	28461	3.0

^{1}H-NMR (D_2O):

Oxidative-acidic depolymerization of acetylated xylan (AcGX)

2.00 g of xylan (MP = 17 KDa) were suspended in 6 mL (3V) of water and the mixture was heated at 70°C under nitrogen flow. 1.0 mL of 30% w/w H_2O_2 (0.33 g of H_2O_2, 9.79 mmol) and 80 µL of 2.5 M H_2SO_4 (0.2 mmol) were added and the reaction mixture was stirred at 70°C for 7.5 h. After 7.5 h IPC revealed Mp equal to 2077 Da (*IPC: 0.5 mL of reaction mixture were neutralized with 0.1M NaOH and 100 µL of 1M BaCl$_2$ were added to precipitate the sulfates. The suspension was filtered and the filtrate was injected in GPC-RI*). The mixture was cooled to room temperature and neutralized with 0.1 M NaOH. The aqueous solution was added drop-wise to 50 mL (25V) of MeOH. The suspension was stirred at room temperature for 1 h and filtered. The solid was washed with 2 x 4 mL of MeOH and dried under vacuum at 50°C for 20 h to yield 1.05 g of depolymerized xylan (y = 53%).

GPC-RI: Mn 1348 Da; Mw 4276 Da; MP 2714 Da; PD 3.12.

^{1}H-NMR (D$_2$O):
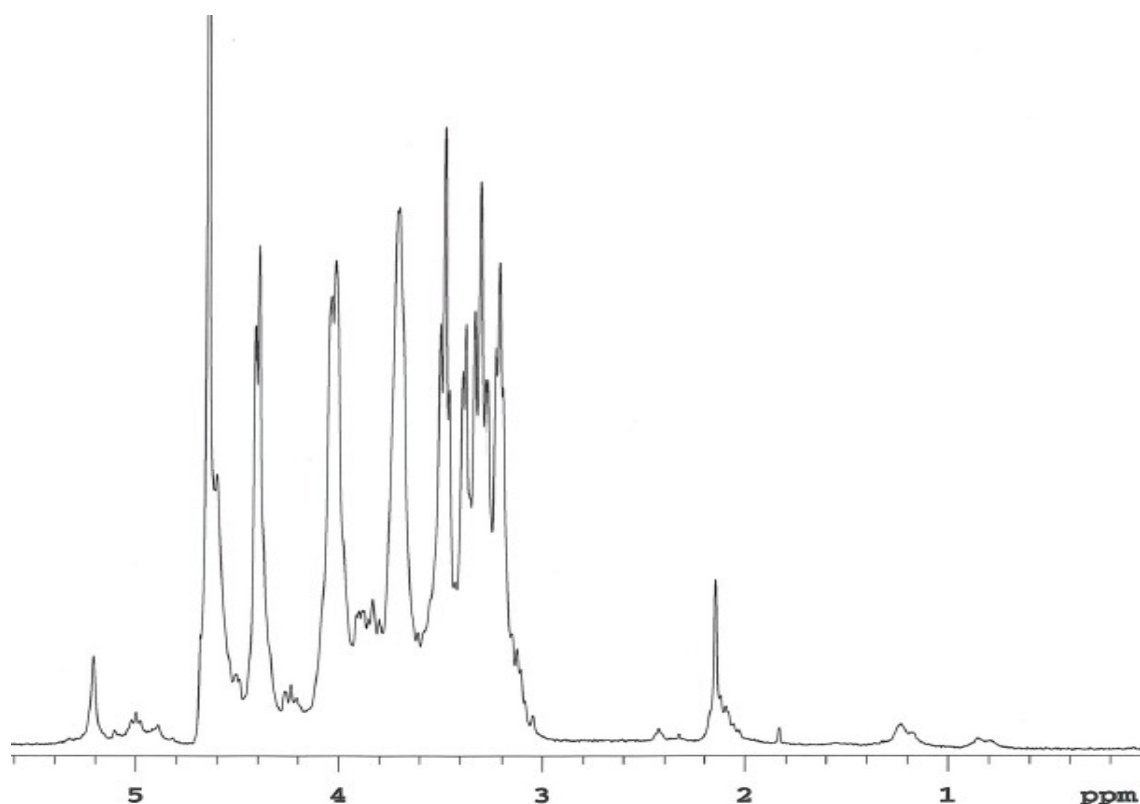
Sulfation of the xylan

In 1 L vessel equipped with two NaOH scrubbers pyridine (840 mL, 825 g, 10.43 mol) was heated at 50°C under nitrogen flow. ClSO$_3$H (180 mL, 315 g, 2.71 mol) was added through a peristaltic pump under the surface of pyridine in 1 h maintaining T at 50-60°C. At the end of the addition the mixture was heated at 60°C for 0.5 h. 100 g of glucuronoxylan (MP 2500 Da) were added and the brownish mixture was heated at 70°C for 2.5 h. The mixture was cooled at 40°C and added dropwise to 5.4 L (54 V) of MeOH. The suspension was stirred at room temperature for 1 h and filtered. The solid was washed with 2 x 200 mL (2 V) of MeOH to yield 350 g of wet pyridinium sulfated polysaccharide. The wet solid was dissolved in 600 mL (6 V) of water at 10°C and 30% NaOH solution was added until pH 12. After 1 h at 10°C, pH was corrected to 6 by addition of AcOH. The aqueous solution was added drop-wise to 6.4 L (64 V) of MeOH. The suspension was stirred at room temperature for 1 h and filtered. The solid was washed with 2 x 200 mL (2 V) of MeOH. The wet solid (400 g) was dried under vacuum at 50°C for 20 h to afford 180 g of crude sulfated polysaccharide.

y (w/w) = 180%.

GPC-RI: Mn 4681 Da; Mw 9015 Da; MP 7081 Da; PD 1.9.

ICP-MS (w/w): S = 18.2%; Na = 13.9%; Ba = 158 ppm.

^{1}H-NMR (D$_2$O):
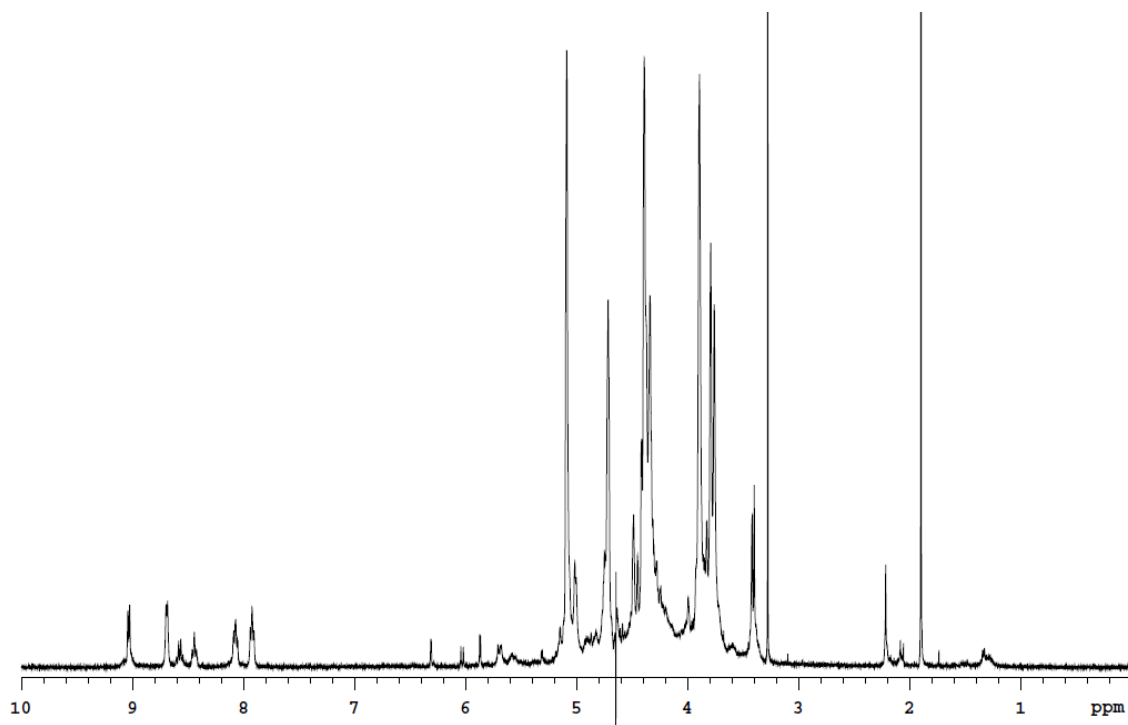
Dialysis of crude sulfated polysaccharide

200 mg of crude sulfated polysaccharide (Mn 3056 Da; Mw 5949 Da; MP 5740 Da; PD 2.2) were dissolved in 200 mL (10 V) of water (conductivity = 6.20 µS/cm) at room temperature and the solution was charged in Slide-A-Lyzer Dialysis Cassette MWCO 2000 Da. The cassette was immerged in 1.8 L of distilled water. The external water was changed after 6 h (conductivity = 18.51 µS/cm) and after 22 h (conductivity = 21.24 µS/cm). After 30 h (conductivity = 9.04 µS/cm) from the starting of the trial the sample solution present in the cassette was freeze dried to yield 140 mg of pure sulfated polysaccharide (recovery = 70%).

GPC-RI: Mn 3362 Da; Mw 5942 Da; MP 5453 Da; PD 1.8.

ICP-MS (w/w): S = 14.9%; Na = 10.9%; Ba = 152 ppm.

^{1}H-NMR (D$_2$O):
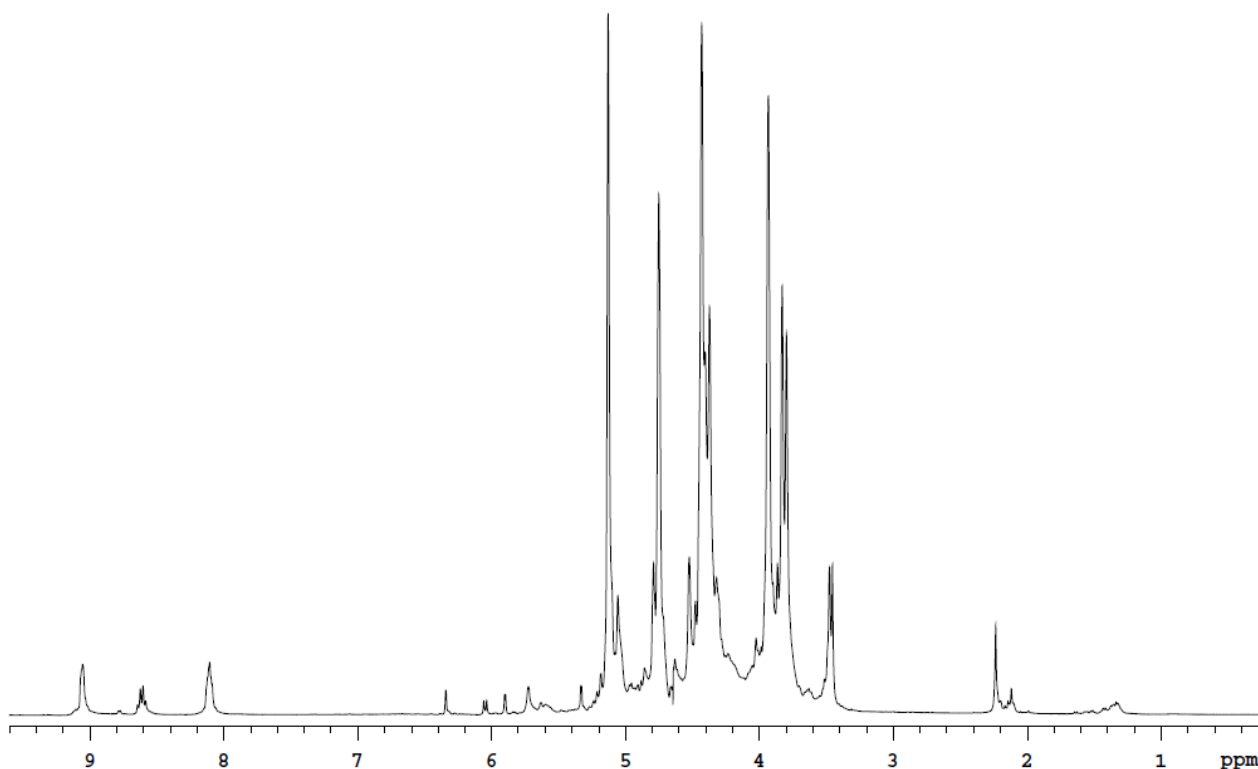
Tangential flow filtration of crude sulfated polysaccharide

1.5 g of crude sulfated polysaccharide (Mn 4793 Da; Mw 7315 Da; MP 5627 Da; PD 1.5) were dissolved in 50 mL (32 V) of water. The solution was purified by tangential flow filtration using a Minimate TFF capsule MWCO 1 KDa. The transmembrane pressure was set to 20 psi and the permeate flow rate was 100 mL/h (J = 0.3 L/ m^2 min). 350 mL of water (7 diafiltration volumes) were added to diafiltrate the solution and then the latter was concentrated to 25 mL (16 V). The obtained solution was freeze dried to yield 820 mg of pure sulfated polysaccharide (recovery = 55 %).

GPC-RI: Mn 4559 Da; Mw 7613 Da; Mp 6019 Da; PD 1.7.

ICP-MS (w/w): S = 15.1%; Na = 10.1%; Ba = 146 ppm.

^{1}H-NMR (D$_2$O):
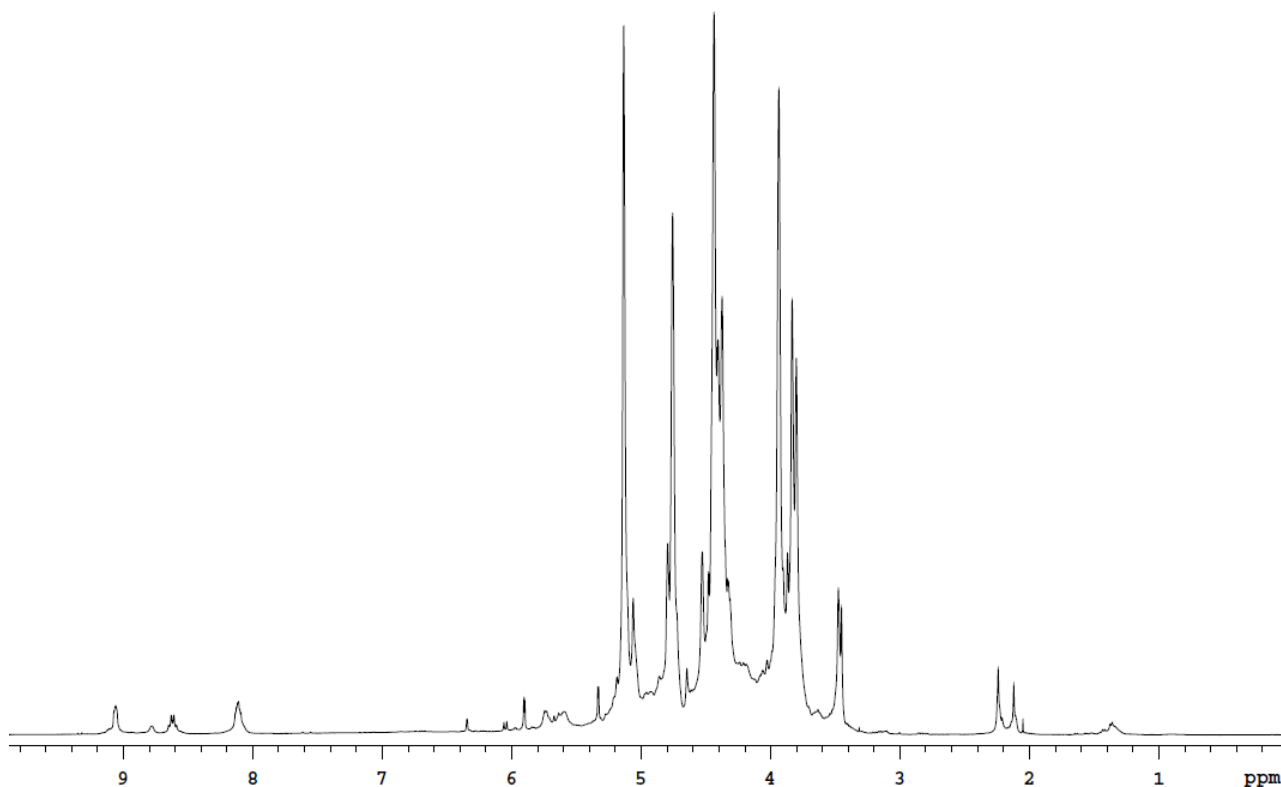
Purification of crude sulfated polysaccharide by precipitation

7.5 g of crude sulfated polysaccharide (Mn 5074 Da; Mw 6636 Da; MP 5586 Da; PD 1.3) were dissolved in 75 mL of water (10 V). 21 mL of 1M BaCl$_2$ (21 mmol) were added and the suspension was stirred at room temperature for 1 h and filtered. The solid was washed with 2 x 15 mL (2 V) of water. The filtrate (105 mL) was charged on 187 mL (25 V) of Amberlyst 15-dry resin conditioned with the following protocol: 375 mL (2 BV) of water, 375 mL (2 BV) of 1M HCl, 375 mL (2 BV) of water, 375 mL (2 BV) of 1M NaOH¸ 375 mL (2 BV) of water. The resin was eluted with water and the eluate containing the product (130 mL) (blue spot in TLC with Cerium Ammonium molybdate reagent) was concentrated under reduced pressure to 15 mL (2 V). The precipitated white crystals were filtered off and the filtrate was diluted to 22.5 mL (3 V) with water. The aqueous solution was added drop-wise to 112.5 mL (15 V) of MeOH. The suspension was stirred at room temperature for 1 h and filtered. The solid was washed with 15 mL (2 V) of 1:5 water: MeOH and with 15 mL (2 V) of MeOH. The wet solid (9.0 g) was dried under vacuum at 50°C for 20 h to yield 4.49 g of pure sulfated polysaccharide (y = 60%).

GPC-RI: Mn 5150 Da; Mw 6776 Da; Mp 5597 Da; PD 1.3.

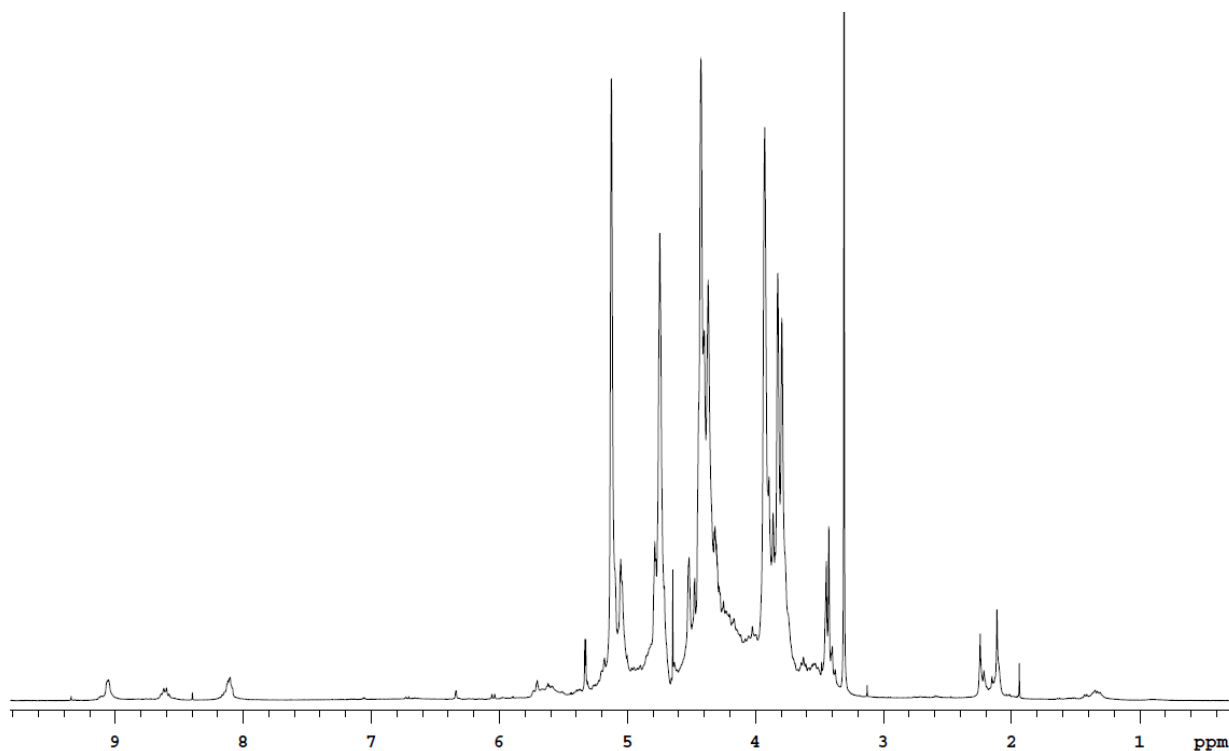ICP-MS (w/w): S = 17.1%; Na = 10.8%; Ba = 4 ppm.

^{1}H-NMR (D$_2$O):
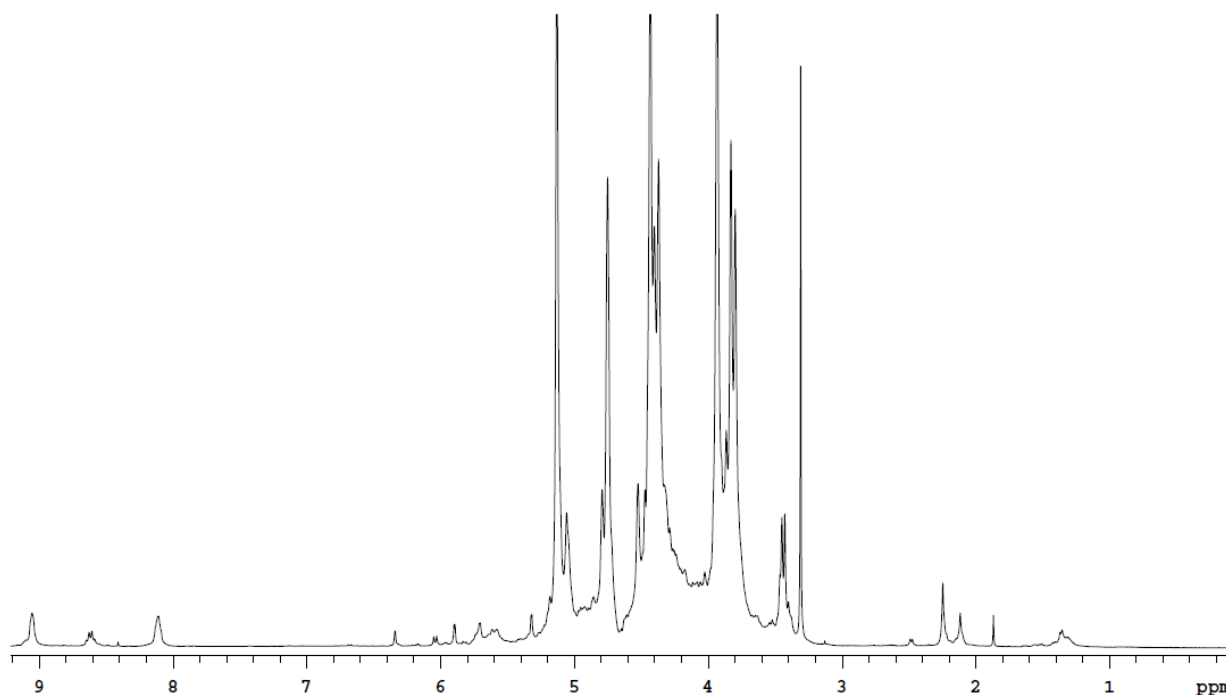
Hydrogen Peroxide bleaching of sulfated polysaccharide

3.00 g of brownish sulfated polysaccharide (Mn 4111 Da; Mw 7527 Da; MP 6356 Da; PD 1.8) were dissolved in 9.0 mL (3 V) of water. 30% w/w H_2O_2 solution (2.4 g, 0.72 g of H_2O_2, 21.1 mmol) was added and the mixture was stirred at 50°C for 8 h. After 8 h, the aqueous solution was added drop-wise to 60.0 mL (20 V) of MeOH. The suspension was stirred at room temperature for 1 h and filtered. The solid was washed with 6 mL (2 V) of 1:5 water: MeOH and with 6 mL (2 V) of MeOH. The wet solid was dried under vacuum at 50°C for 20 h to yield 2.60 g of pale yellow sulfated polysaccharide (y = 87%).

GPC-RI: Mn 3882 Da; Mw 7109 Da; Mp 6095 Da; PD 1.8.

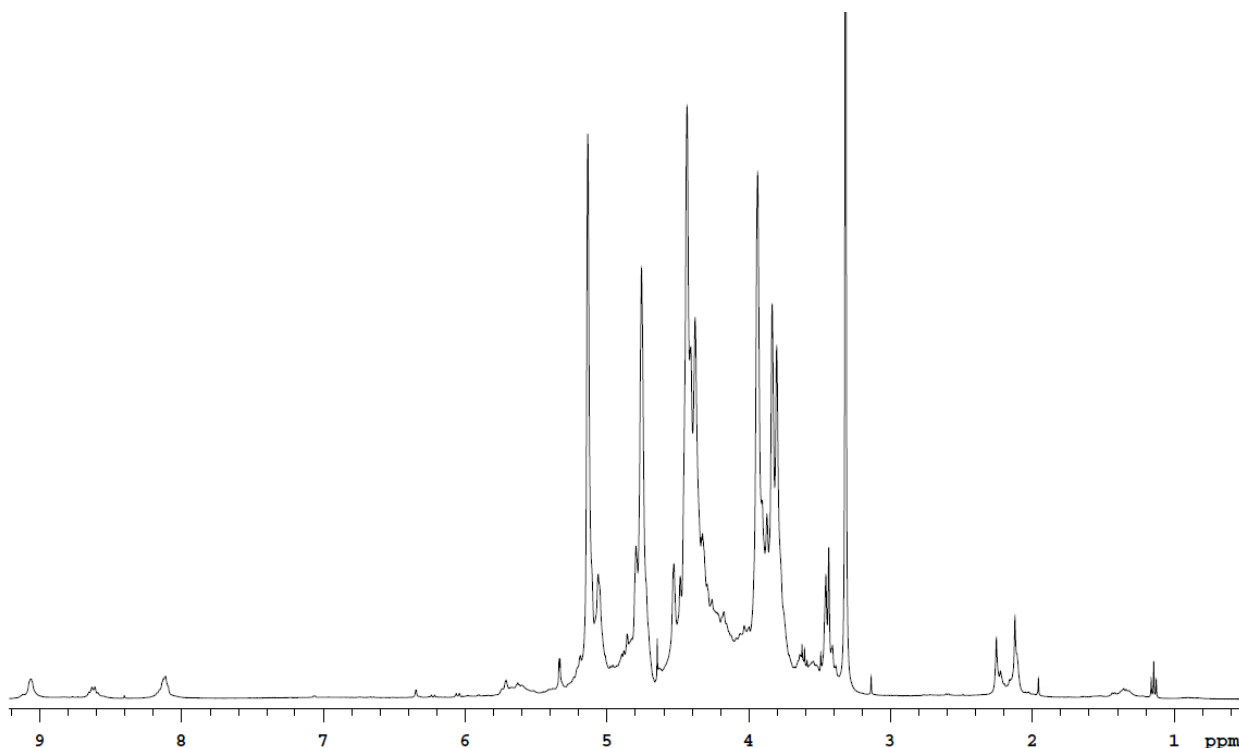ICP-MS (w/w): S = 15.3%; Na = 12.7%; Ba = 17 ppm.

^{1}H-NMR (D_2O):

Chlorine Dioxide bleaching of sulfated polysaccharide

3.00 g of brownish sulfated polysaccharide (Mn 4111 Da; Mw 7527 Da; MP 6356 Da; PD 1.8) were dissolved in 9.0 mL (3 V) of water. 80% $NaClO_2$ (112 mg, 90 mg of $NaClO_2$, 0.99 mmol) and AcOH (46 mg, 0.77 mmol) were added and the mixture was stirred at room temperature for 4 h. After 4 h, the aqueous solution was added drop-wise to 60.0 mL (20 V) of MeOH. The suspension was stirred at room temperature for 1 h and filtered. The solid was washed with 6 mL (2 V) of 1:5 water: MeOH and with 6 mL (2 V) of MeOH. The wet solid was dried under vacuum at 50°C for 20 h to yield 2.60 g of yellowish sulfated polysaccharide (y = 87%).

GPC-RI: Mn 3798 Da; Mw 7112 Da; Mp 6194 Da; PD 1.9.

ICP-MS (w/w): S = 15.9%; Na = 12.8%; Ba = 13 ppm.

^{1}H-NMR (D_2O):

Sodium Hypochlorite bleaching of sulfated polysaccharide

3.00 g of brownish sulfated polysaccharide (Mn 4111 Da; Mw 7527 Da; MP 6356 Da; PD 1.8) were dissolved in 9.0 mL (3 V) of water. 13% NaClO solution (3.0 mL, 0.39 g of NaClO, 5.3 mmol) was added and the mixture was stirred at room temperature for 1 h. After 1 h, the aqueous solution was added drop-wise to 60.0 mL (20 V) of MeOH. The suspension was stirred at room temperature for 1 h and filtered. The solid was washed with 6 mL (2 V) of 1:5 water: MeOH and with 6 mL (2 V) of MeOH. The wet solid was dried under vaccum at 50°C for 20 h to yield 2.55 g of off-white sulfated polysaccharide (y = 85%).

GPC-RI: Mn 3717 Da; Mw 7247 Da; Mp 6228 Da; PD 2.0.

ICP-MS (w/w): S = 15.8%; Na = 13.1%; Ba = 13 ppm.

^{1}H-NMR (D$_2$O):
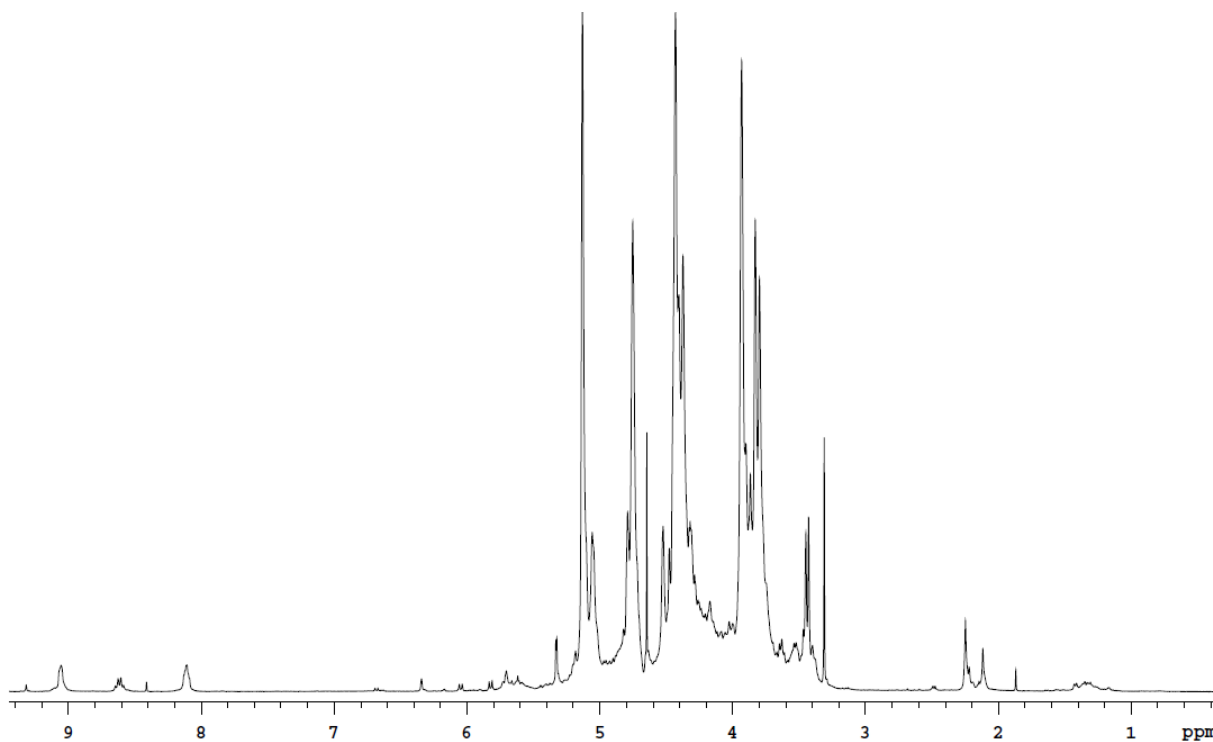
References

[1] C.F: Shen, *Ph. D. dissertation, The City University of New York,* **1992**.

[2] P. Manos, M. Standford, *Int. J. Plant Sci.*, **2001**, 162, S77-S93.

[3] G. Caudullo, E. Welk, J. San-Miguel-Ayanz, *Data Brief*, **2017**, 12, 662-666.

[4] K. Kosikova, M. Hricovini, C. Cosentino, *Wood Sci. Technol.*, **1999**, 33, 373-380.

[5] K. Freudenberg, E. Braun, *Liebigs Ann.*, **1928**, 5, 288-304.

[6] T. Timell, *Adv. Carbohydr. Chem. Biochem.*, **1967**, 19, 247-302.

[7] S. Perez, K. Mazeau, *Polysaccharides, structural diversity and functional versatility*, Crcpress, **2005**, 2, 41-68.

[8] E. Sjostrom, *Wood Chemistry. Fundamentals and applications,* Academic Press, **1993**, 4, 71-89.

[9] E. Jaako Vilen, *Ph. D. dissertation, University of Toronto,* **1999.**

[10] J. Jones, L. Wise, *J. Chem. Soc.*, **1952**, 2570, 3389-3393.

[11] A. Teleman, M. Tenkamen, A. Jacobs, O. Dahlman, *Carbohydr. Res.*, **2002**, 337, 373-377.

[12] J. Bromeley, M. Busse-Wicher, T. Tryfona, J. Mortimer, Z. Zhang, D. Brown, P. Dupree, *Plant J.*, **2013**, 74, 423-434.

[13] M. Busse-Wicher, T. Gomes, T. Tryfona, N. Nikolovski, K. Stott, N. Grantham, D. Bolam, P. Dupree, *Plant J.*, **2014**, 79, 1-14.

[14] E. Husemann, *J. Prakt. Chem.*, **1940**, 155, 13-64.

[15] E. Sjostrom, B. Enstrom, *Tappi J.*, **1967**, 50, 32-36.

[16] A. Teleman, P. Larsson, T. Iversen, *Cellulose*, **2001**, 8, 209–215.

[17] J. Gierer, *Wood Sci. Technol.*, **1986**, 20. 1-33.

[18] C. Brage, T. Erikson, J. Gierer, *Holzforschung*, **1991**, 45, 23-30.

[19] J. Kolar, B. Lindgren, *Acta Chem. Scand.*, **1982**, 36, 599-605.

[20] Bretherick's Handbook of Reactive Chemical Hazards, Elsevier, 1443-1445.

[21] C. Young, *IUPAC Solubility data series*, 12, 454-456.

[22] G. Ferweda, *Pulp Pap. Can.*, **2004**, 12, 63-66.

[23] E. Hagglund, B. Lindberg, J. McPherson, *Acta Chem. Scand.*, **1956**, 10, 1160-1164.

[24] J. Rowley, S. Decker, W. Michener, S. Black, *3 Biotech*, **2013**, 3, 433-438.

[25] I. McDonald, *J. Chem Soc.*, **1952**, 3183-3186.

[26] I. Uhliarikova, M. Vranska, B. McCleary, P. Biely, *Biochim. Biophys. Acta*, **2013**, 3365-3372.

[27] J. Farrand, D. Johnson, *J. Org. Chem.*, **1976**, 36, 3606-3612.

[28] W. Lawrence, R. McKelvey, D. Johnson, *Sven. Papperstidn.*, **1980**, 83, 11-18.

[29] V. Goncalves, D. Evtuguin, M. Domingues, *Carbohydr. Res.*, **2008**, 343, 256-266.

[30] S. Rahman, N. Sugitani, M. Hatsu, K.Takamizawa, *Can. J. Microbiol.*, **2003**, 49, 58-64.

[31] V. Sivova, G. Nosalova, L. Jurecek, J. Turjan, S. Vickova, P. Bystricky, D. Kantarova, P. Capek, *Starch*, **2015**, 67, 1-8.

[32] A. Sarbu, F. Goncalves, M Pinho, *Carbohydr. Polym.,* **2003**, 53, 1-8.

[33] E. Lattova A. Ebringerova, R. Toman, M. Kacurakova, *Chem. Pap.*, **1992**, 46, 66-69.

[34] E. Husseman, O. Westphal, B. Kickhofen, **1954**, patent US2689848.

[35]J. Raman, B. Pillai, S. Balaji, H. Patel, D. Dhameliya, G. Jinjala, **2012**, patent appl. WO2012/101544.

[36] G. Garrote, H. Domiguez, J. Parajo, *J. Chem. Technol. Biotechnol.*, **1999**, 74, 1101-1109.

[37] B. Kuznetsov, A. Efremov, V. Levdanskii, A. Kuznetsova, N. Polezhayveva, T. Shilkina, I. Krotova, *Bioresour. Technol.,* **1996**, 58, 181-188.

[38] M. Maloney, T. Chapman, *Biotechnol. Bioeng.*, **1985**, 27,355-361.

[39] V. Stajic, N. Cheetham, A. Bell, **2015**, patent US9120877.

[40] E. Gilbert, *Chem. Rev.,* **1962**, 62, 549-589.

[41] L. Schwartz, Introduction to Tangential Flow Filtration for Laboratory and Process Development Applications, *Pall corporation*.

[42] S. Collier, P. Ghosh, *Ann. Rheum. Dis.*, **1989**, 48, 372-381.

# 4.Evaluation of products similarity and Conclusions

In the second chapter several analytical methods for the characterization of the sulfated polysaccharide (SP) were disclosed and they include Gel Permeation Chromatography, Nuclear magnetic resonance, determination of sulfate content (conductometric titration, ICP-MS and ionic chromatography), Capillary Zone Electrophoresis and Liquid Chromatography - Mass Spectrometry. In the third chapter we discussed the key steps of the process for the preparation of SP from beech wood. In this chapter we will compare the Originator API and the product obtained with the process described in chapter 3. Indeed, the aim of the project is the introduction of the generic version of the sulfated polysaccharide and hence the sameness with the originator's product is one of the main requirements. As already discussed, because the sulfated polysaccharide is a complex molecule, the comparison can not be based on a single analysis, but requires orthogonal analytical methods, as recommended by Food and Drug Administration.

4.1 Gel Permeation Chromatography (GPC)

As described in chapter 2, Gel Permeation Chromatography (GPC) allows to study the molecular weight distribution, which is a key feature of a polymer. The analysis was performed both with the Refractive Index (RI) detector, which affords results relative to a calibration curve obtained with Pullulan standards, and with the Multi Angle Laser Light Scattering (MALLS) detector, which provides absolute results. GPC-RI has a low reproducibility between analytical sessions run in different days or weeks, because small differences in eluent composition and other experimental conditions affect the results significantly. Therefore, two samples under comparison must be analyzed in the same session. In Table 4.1 the results of molecular weight distribution of a sample of Originator and Indena SP obtained in the same session of analysis are reported. The molecular weight values (Mn and Mw) of the two products are comparable, but the polydispersity of Indena SP is slightly lower. The same conclusion can be extracted from the overlay of the chromatograms (Figure 4.1) which show two similar distribution, but the curve of Indena SP is slightly narrower than the originator's one.

Sample	Mn (Da)	Mw (Da)	PD
Originator SP	2125	6471	3.0
Indena SP	2344	6126	2.6

Table 4.1. GPC-RI results of originator and Indena SP
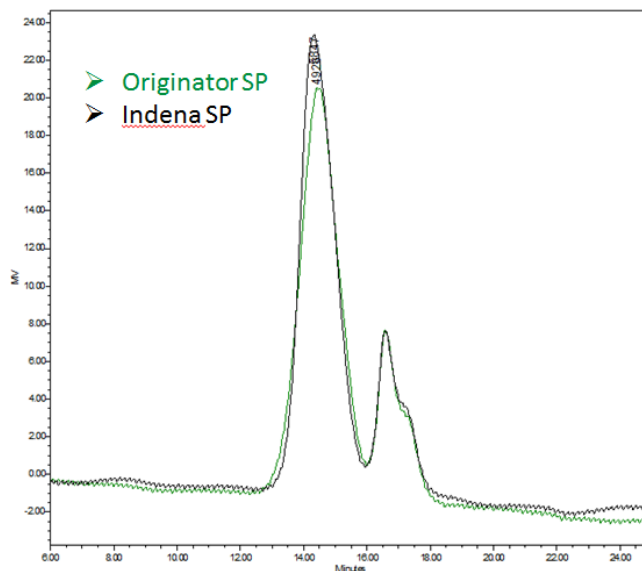
Figure 4.1. Overlay of GPC-RI chromatograms of originator and Indena SP

GPC-MALLS has higher inter-session reproducibility and affords more reliable results. However, it is crucial to choose the appropriate value of all the constants (e.g. dn/dc) which correlate the light scattered with the molecular weight of the polymer. Table 4.2 shows the results obtained for Originator SP and Indena SP. Indena SP's Mn is 400 Da higher and its Mw 400 Da lower than those of the commercial API. This means that our version of the sulfated polysaccharide has a molecular weight distribution centered on the commercial product, but less polydisperse.

Sample	Mn (Da)	Mw (Da)	PD
Originator SP	4355	5621	1.3
Indena SP	4937	5234	1.1

Table 4.2. GPC-MALLS results of originator and Indena SP

In both cases the comparison was carried out on a single batch of Originator and a single batch of Indena SP. Obviously, there will be some variability among the different lots of both products. Therefore, in the future a statistical evaluation will be more appropriate for a full comparison.

Anyway, from GPC-RI and GPC-MALLs analysis it is clear that our version of the sulfated polysaccharide has a molecular weight comparable to the Originator product, with a slight difference only in the polydispersity.

4.2 Nuclear magnetic Resonance (NMR)

Figure 4.2 shows the overlaid proton spectra of Indena SP (blue) and Originator SP (red). The visual comparison shows high similarity between the two products that share all the main signals of sulfated xylose units in the central region between 5 and 3.5 ppm. Moreover, in Indena SP there are also other key features of Originator compound like the acetyl groups (methyl groups at 2 ppm) and the presence of a few rhamnose units (methyl groups at 1.2 ppm). In the aromatic protons area (8-9 ppm) both spectra feature the signals of pyridinium groups attached to the anomeric position of the reducing end of xylose units.
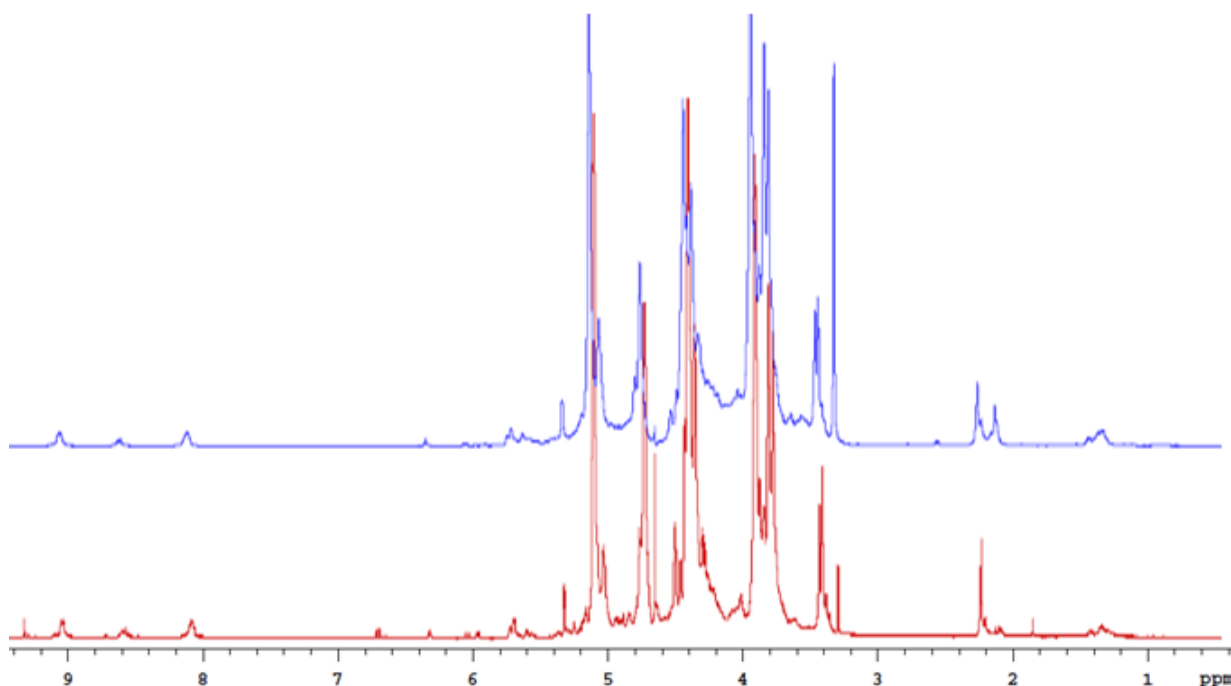
Figure 4.2. Overlay of ^{1}H-NMR spectra of Indena SP (blue) and Originator SP (red).

More insight can be attained by the ratios between the integrals of the different entities present in the sulfated polysaccharide, as described in paragraph 2.2b. In Table 4.3 for each integral ratio on one side we reported the minimum and maximum values found among the different batches of Originator product and on the other side the value of a typical Indena sulfated polysaccharide. As far as acetyl groups are concerned, the two products can be considered similar, because the content of acetyl groups in position 3 of xylose units decorated with methylglucuronic acid (Entries 1 and 2) is almost the same, while the overall degree of acetylation (Entry 3) is slightly higher in the Indena product. The content of 4-O-methylglucuronic acid (MGA) (Entries 4 and 5) of Indena SP is within the ranges of Originator SP regardless if the ratios were calculated on anomeric protons or on methyl groups. The average degree of polymerization (Entry 6) calculated by NMR

confirms GPC results and the similarity of chain length of the two versions of the sulfated polysaccharide. Likewise, the relative amount of pyridinium moieties attached to the anomeric position of terminal xylose (Entry 7) is comparable in both the products.

Entry	Ratio	Formula	Originator Min	Originator Max	Indena SP
1	3-*O*-acetylation	$$\dfrac{CH_3\ 3OAc}{OCH_3MGA\ (xyl\ Ac)\ and\ OCH_3\ MGA\ (Xyl)}$$	0.31	0.41	0.30
2	3-*O*-acetylation bis	$$\dfrac{OCH_3MGA\ (xyl\ Ac)}{OCH_3MGA\ (xyl\ Ac)\ +\ OCH_3\ MGA\ (Xyl)}$$	0.42	0.58	0.47
3	Overall degree of acetylation	$$\dfrac{(CH_3\ all\ Ac)/3}{sum\ of\ anomeric\ xyl}$$	4.29 %	5.99 %	6.87 %
4	MGA amount	$$\dfrac{sum\ of\ anomeric\ MGA}{sum\ of\ anomeric\ xyl}$$	5.39 %	8.26 %	5.96 %
5	MGA amount bis	$$\dfrac{(OCH_3MGA\ (xyl\ Ac)\ +\ OCH_3\ MGA\ (Xyl))/3}{sum\ of\ anomeric\ xyl}$$	7.19 %	8.76 %	7.96 %
6	Average degree of polymerization	$$\dfrac{sum\ of\ anomeric\ xyl}{sum\ of\ anomeric\ RE\ xyl}$$	8.56 %	7.56 %	7.30 %
7	Pyridinium	$$\dfrac{H4\ Py}{sum\ of\ anomeric\ xyl}$$	2.07 %	2.96 %	2.11 %

Table 4.3. Comparison of ^{1}H-NMR integral ratios of Originator and Indena SP. Sum of anomeric xyl = xyl RE α *100/35 + xyl int/NR +(CH$_3$ 3-OAc)/3. RE = reducing end, NR = not reducing end.

In conclusion, within the error in the integration of complex spectra, we can consider Originator and Indena SP comparable from a qualitative and quantitative point of view.

4.3 Content of sulfates

As discussed in paragraph 2.3, this parameter can be determined by three different methods: conductimetric titration, Inductively coupled Plasma Mass spectrometry (ICP-MS) and ionic chromatography. ICP-MS allows to determine also the amount of sodium, which is present as counter-ion of sulfates, and barium, which can come from the process. On the other hand, ionic chromatography is the only technique able to distinguish between sulfates bound to the xylose

chains and free sulfates present as inorganic impurities. Table 4.4 shows a summary of the results obtained on a lot of Originator SP and Indena SP with the three analytical techniques. The content of sulfates is expressed as percentage of sulfur in the product. The results are slightly different depending on the analytical technique (e.g. Originator SP has 15.5% of sulfur by titration and 16.8% by Ionic chromatography). However, with all three methods both the products have sulfur in the range between 15 and 17% and hence the content of sulfates in Originator and Indena SPs can be considered comparable. Moreover, according to ICP-MS data sodium is present in similar amount in the two SPs and barium has very low concentrations (ppm units). Finally, ionic chromatography reveals that free sulfur in Indena SP (0.07%) is lower than in Originator SP (0.20%).

Sample	%S by Titration	ICP-MS			Ionic chromatography	
		% S	%Na	Ba (ppm)	% Bound S	% Free S
Originator SP	16.6	15.5	12.0	3	16.8	0.20
Indena SP	16.2	17.1	10.8	4	15.7	0.07

Table 4.4. Comparison of results obtained by conductimetric titration, ICP-MS and Ionic chromatography of Originator and Indena SP.

4.4 Capillary Zone Electrophoresis (CZE)

As explained in paragraph 2.4, capillary zone electrophoresis (CZE) separates the different species of the sulfated polysaccharide according to their ability to migrate under the influence of an electrical field. The analysis affords a typical fingerprint which is crucial to establish similarity. In Figure 4.3 the overlay of typical CZE profile of Indena and Originator products are shown
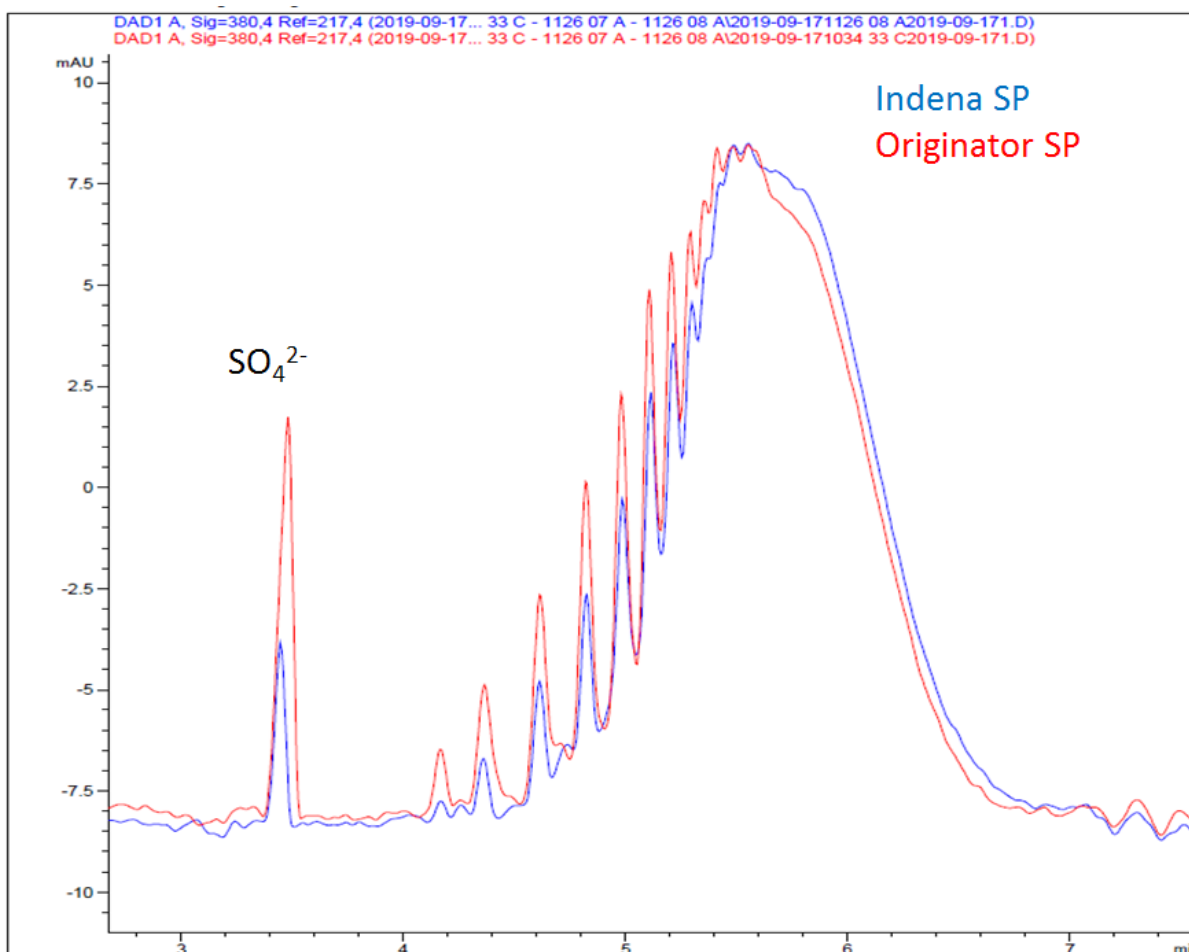
Figure 4.3. Overlay of electropherograms of Indena SP (blue) and Originator SP (red).

In the first part of the electropherogram at 3.4 min we can observe a main peak related to the presence of inorganic sulfates. The sulfates in Indena SP are lower than in Originator SP, as shown also by determination of free sulfates by ionic chromatography (paragraph 4.3). At longer migration time the curve of the sulfated polysaccharide appears. First of all, we can observe that the width (between 4 and 7 min of migration time) and the height of the curve are almost identical. Moreover, both samples show peaks which emerge from the curve and they have exactly the same migration times in the two samples. The difference between the two profiles is the intensity of the peaks which are slightly more intense in Originator SP than in Indena SP. However, CZE is not a quantitative analysis and it is likely that small differences in the composition of the sulfated polysaccharide can generate visible differences in the intensity of the peaks. Concluding, in Indena and Originator sulfated polysaccharide the shape of the curve and the migration times of the peaks are really similar and hence CZE fingerprints result comparable

4.5 Liquid Chromatography-Mass spectrometry (LC-MS)

Ion-pair reverse phase (IRRP) chromatography coupled to electrospray ionization (ESI) mass spectrometry provides not only a typical fingerprint, but also structural information of the species present in the SP.

Figure 4.4 shows the comparison between LC-MS profiles of a sample of Indena SP (black) and Originator SP (brown) recorded in the same analytical session.
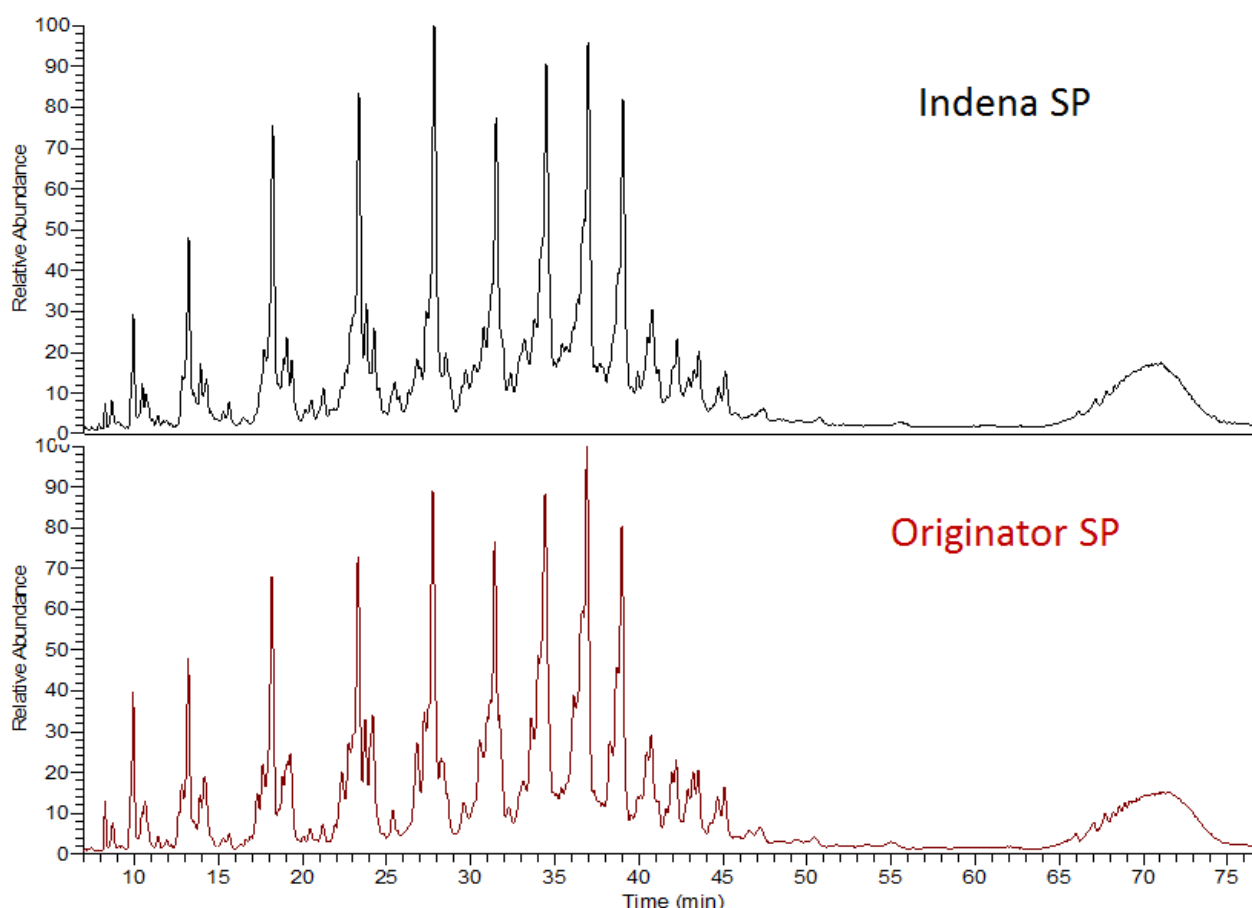

Figure 4.4. Comparison of LC-MS profile of Indena SP and Originator SP.

The two chromatograms are really similar with the typical fingerprint characterized by a group of peaks repeating at different retention times. Moreover, in both products you can observe the broad peak at RT = 70 min related to the presence of high MW oligomers, which are not efficiently separated by the chromatographic system. A deeper evaluation of the similarity can be achieved by comparison of the extracted ion currents (EICs) of the species present in the sulfated polysaccharide (Figure 4.5). We looked for the following oligomers: xylose chain completely sulfated but on the anomeric position (Xyl_nS_{2n+1}); xylose chain completely sulfated with pyridinium moiety attached to the anomeric position ($Xyl_nS_{2n+1}Pyr_1$); xylose chain completely sulfated with double bond on one unit ($Xyl_nS_{2n+1}-H_2O$); xylose chain completely sulfated having as the terminal

unit an open ring aldehyde conjugated to double bond between position 2 and 3 (Xyl_nS_{2n}-H_2O); xylose chain bearing 4-O-methylglucuronic acid ($Xyl_nMGA_1S_{2n+2}$); xylose chain bearing 4-O-methylglucuronic acid and an acetyl group, probably in position 3 of xylose bearing MGA residue ($Xyl_nMGA_1S_{2n+1}Ac_1$); xylose chain sulfated with an additional methyl group probably related to the presence of a Rhamnose unit ($Xyl_nS_{2n+1}Me_1$).

First of all, in Figure 4.5 we can observe that all the above-mentioned species are present both In Originator and Indena SP. Moreover, the extracted ion currents are almost identical in the two products and hence the distribution of the species among the different degree of polymerization is almost identical. Without reference standards LC-MS does not allow to quantify the amount of a defined species in the product so that a quantitative comparison is not possible.
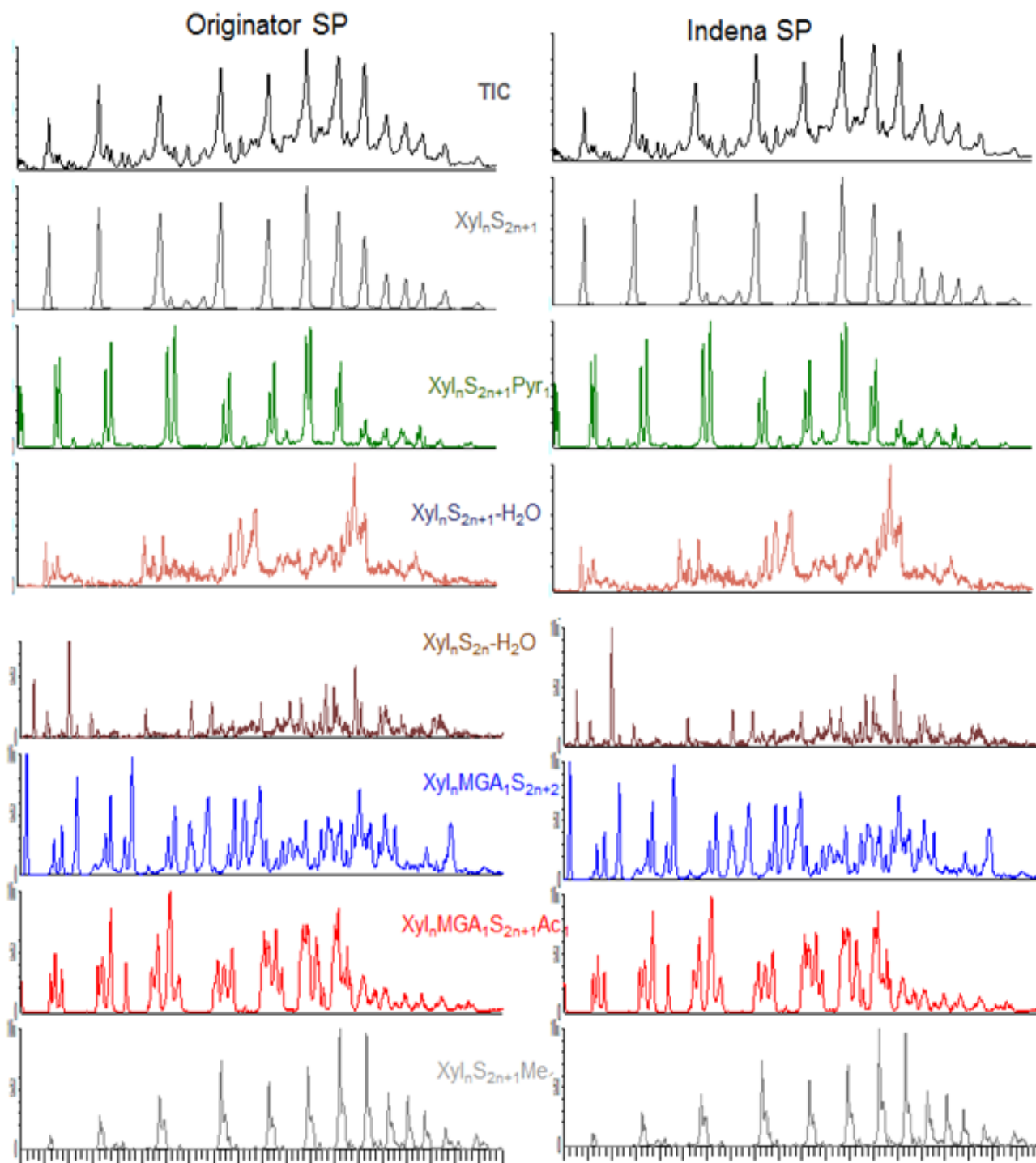
Figure 4.5. Comparison of extracted ion currents of Originator and Indena SP.

Considering the details of the sulfate polysaccharide composition shown by LC-MS analysis, we can conclude that the Indena product is very similar to the Originator one. Indeed, the two products have not only comparable Total Ion Current (TIC), but also similar Extracted Ion Currents (EICs) of the main species present in the SP.

207

4.6 Infrared Spectroscopy

This technique, as discussed in paragraph 2.6, is not very sensitive and does not show important differences between different versions of the sulfated polysaccharide. For example, substandard SP (e.g. the Indian one) which is characterized by the complete absence of acetyl groups, has an IR spectrum almost identical to Originator SP. Anyway, Food and Drug Administration (FDA) recommends the use of infrared spectroscopy to assess the sameness between two complex APIs. Therefore, we performed IR analysis on Indena and Originator SP and the two spectra are reported in Figure 4.6.
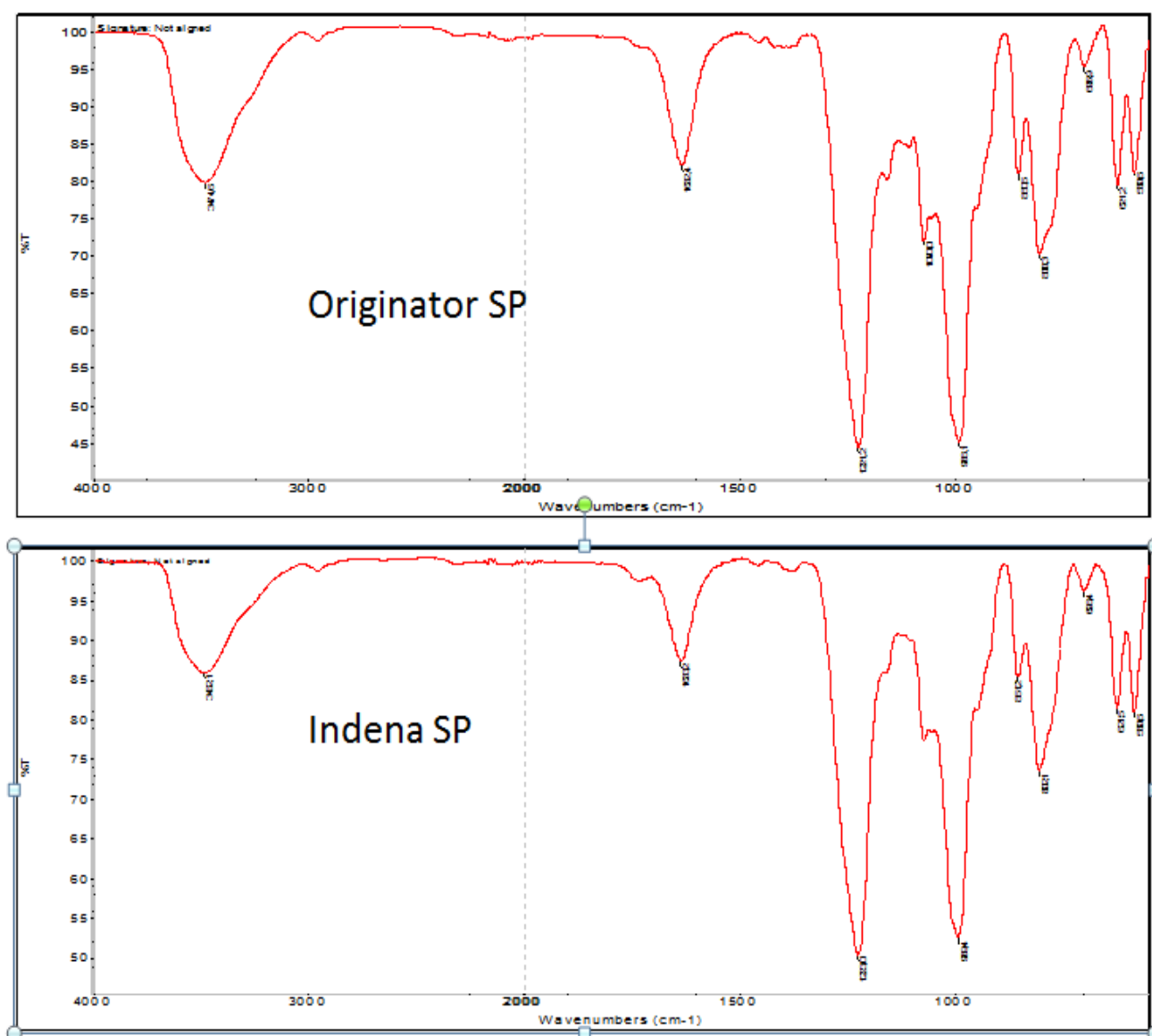

**Figure 4.6. Infrared spectra of Originator and Indena SP.**

The two spectra appear very similar and have the same vibrational bands.

4.7 Conclusions

A comparison between the Originator product and our version of the sulfated polysaccharide (SP) can be achieved by several orthogonal analytical methods.

First of all, Originator and Indena SP have comparable molecular weight, according to Gel Permeation Chromatography with both Refractive Index and Multi Angle Laser Light Scattering detectors. The similarity of the chain lengths of the two polymers is confirmed also by the ratio between the integrals of anomeric protons signals in ^{1}H-NMR spectrum which allows to calculate the average degree of polymerization. Furthermore, Originator and Indena versions of the sulfated polysaccharide have a comparable degree of sulfation according to three different analytical techniques (conductimetric titration, inductively coupled plasma mass spectrometry and ionic chromatography).

Both products display similar modifications of the sulfated xylose chains (Figure 4.7). Indeed, the xylose chain completely sulfated but at the anomeric position, **4**, is the most abundant species, but there are other species which characterize the sulfated polysaccharide. For example, at the anomeric position the hydroxyl group can be replaced by pyridinium moieties, generating structure **3**. A xylose unit can also have a double bond in a completely sulfated chain, species **5**, or coupled to an open ring aldehyde, species **2**. Furthermore, the xylose chain can have a ramification of 4-*O*-methylglucuronic acid (MGA) and the position 3 of xylose bearing MGA residue can be acetylated like in structure **7**. Finally, xylose is not the only sugar present in the product, but there are oligomers **8** characterized by the presence of a Rhamnose unit. The presence of these species can be revealed by extracting the ion currents in LC-MS and studying a few characteristic signals (e.g. aromatic proton area, acetyl groups at about 2 ppm) of ^{1}H-NMR spectrum. These analysis show that all the above-mentioned species are present in both the Originator and Indena version of the sulfated polysaccharide.
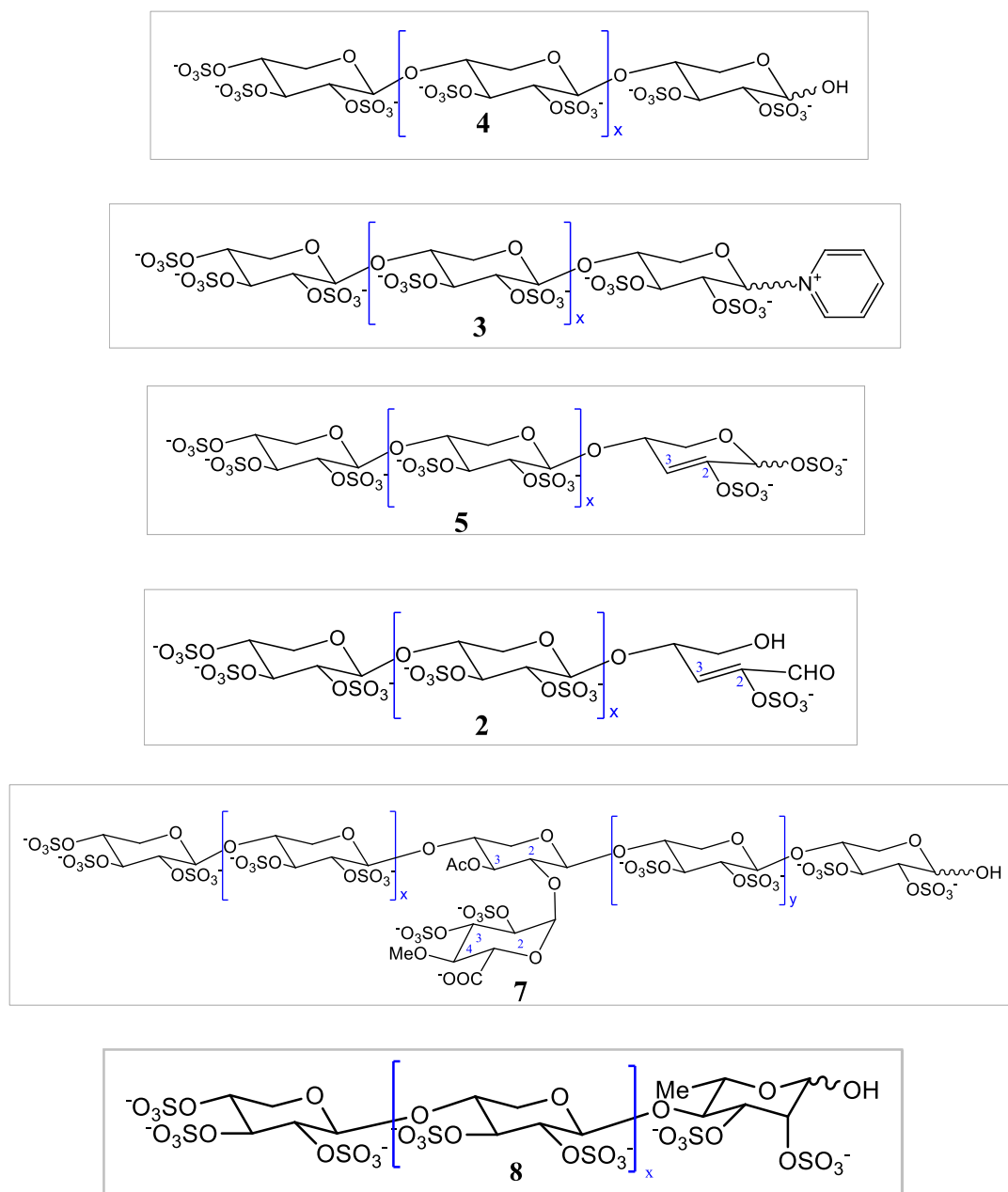
Figure 4.7. Summary of the species present in Originator and Indena sulfated polysaccharide.

Finally, the similarity between the two products is underlined by profiles in CZE, where the general shapes of the curves are comparable and the migration times of the most intense peaks are the same. Very similar fingerprints are also obtained with Total Ion Currents (TICs) in LC-MS.

In conclusion, we designed a process which allows to prepare from beech wood a version of the sulfated polysaccharide (Indena SP) very similar and comparable to Originator SP according to several orthogonal analytical methods. Therefore, our version is a good candidate to be the first generic version of the sulfated polysaccharide in the United States approved by Food and Drug Administration (FDA).