Exploring the Structure-Polymorphism Relationship of the ROY Polymorphophore via the use of Solid-Form Informatics

Samuel Lee Meredith

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of Leeds

School of Chemical and Process Engineering

200874867

November 2024

The candidate confirms that the work submitted is his own, and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I first thank Martin, Colin and the rest of the Fishwick/McPhillie group for their support. I thank Bao and Robert for their key insight and guidance during the earlier stages of the project. I'd also like to thank Andy Maloney, without whom the computational aspect of this project would have been impossible.

I'd like to thank all who have helped with the collection of data, particularly Chris Pask (for being my personal helpline for all things crystallography), Jeanine Williams (for the endless chromatography samples).

I'd like to extend my thanks to the Fishwick/McPhillie and Bon/Wilson groups, in particular Scott Grossman (for waffling with me about all things science) and Aidan Johnson (a partner for pints, and a wonderful friend).

I'd also like to extend my thanks to the Molecules to Product CDT, particularly Angela Morrison and Elaine Martin for their unwavering support and enthusiasm.

Thank you to the CCDC, Pfizer and Syngenta for their support and insight.

Finally, I'd like to thank my mum, my granny and grandad and my little sister, who always believed in me and have made the whole thing possible.

List of Figures

Abstract

Polymorphism is a phenomenon by which a single chemical entity can crystallise in multiple distinct crystalline forms, each displaying unique physicochemical properties. As such, polymorphism is a relevant issue across any field which utilises crystalline forms, in particular the pharmaceutical industry. Significant time and money is spent industrially to explore the polymorphic landscape of candidate molecules, using methods such as High-Throughput Screening and Crystal Structure Prediction. Despite its relevance, little progress has been made towards any understanding of links between chemical structure and polymorphism.

A polymorphophore is a class of compounds which, analogous to the concept of a pharmacophore, has the ability to impart structures with polymorphism when incorporated into a new chemical structure. These structures could provide a link between chemical structure and the phenomenon of polymorphism, but to date only two polymorphophore structures have been confirmed and investigated experimentally (N-phenylbenzamide and Fenamates). 5-methyl-2-[(2-nitropphenyl)amino]-3-thiophenecarbonitrile (aka ROY) is a compound for which 12 unique single component crystal forms have been discovered, which has also been suspected of being a polymorphophore but has yet to be confirmed as such.

In this work the ROY structure was investigated using Solid Form Informatics (SFI), an informatics workflow which utilises the Cambridge Structural Database (CSD), a public library of crystal information containing over 1.25 million structures. SFI utilises informatics tools developed from the data within the CSD to explore the intra-, inter- and supramolecular environments of the 12 ROY polymorphs, in order to find structural features which may influence the polymorphism of ROY. These structural features were used to design and synthesise a set of 12 structural analogues of ROY, which were then crystallised and analysed using SFI to track the changes in intra-, inter- and supramolecular environment introduced by changes in structure.

Finally, these 12 structural analogues were subject a low-/medium throughput polymorph screen, during which nine analogues were confirmed to display at least two polymorphic forms. These data confirm the classification of ROY as a polymorphophore. However, the lack of data regarding the properties, behaviour and influence of polymorphophores in general make it difficult to provide a full evaluation of the implications of this classification.

List of Abbreviations

4’-methyl-ROY	5-methyl-2-[(2-nitro-4-methyl)amino]-3-thiophenecarbonitrile
5-acetyl-ROY	5-acetyl-2-[(2-nitro-4-methyl)amino]-3-thiophenecarbonitrile
AIDS	Autoimmune disease syndrome
Amino-ROY	5-methyl-2-[(2-aminopehnyl)amino]-3-thiophenecarbonitrile
Br-ROY	Brominated ROY-analogue
CCDC	Cambridge Crystallographic Data Centre
CIF	Crystal Information File
CSD	Cambridge Structural Database
CSP	Crystal Structure Prediction
DCM	Dichloromethane
Demethyl-ROY	2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile
DFT	Density Functional Theory
ΔH_{lat}	Enthalpy of lattice formation
DIBAL-H	Diisobutylaluminium hydride
DMF	Dimethylformamide
DSC	Differential Scanning Calorimetry
ΔU_{inter}	Intermolecular interaction energy
ΔU_{lat}	Internal lattice energy
El-ROY	Ethynyl linked ROY derivative
ENaCt	Encapsulated Nanocrystallisation
EtOAc	Ethyl acetate
EtOH	Ethanol
FA	Fenamic acid
Fenamate	2-anilinobenzoic acid
FIM	Full Interaction Map
Fu-ROY	5-methyl-2-[(2-nitrophenyl)amino]-3-furancarbonitrile
H_2	Hydrogen gas
H-bond	Hydrogen bond

HBP	Hydrogen Bond Propensity
HCl	Hydrochloric acid
HPLC	High Performance Liquid Chromatography
HRMS	High-Res Mass Spectrometry
InChI	International Chemical Identifier
KOH	Potassium hydroxide
LiCl	Lithium chloride
LogP	Partition coefficient
L-proline	(2S)-pyrrolidine-2-carboxylic acid
MA	2-[(2,3-dimethylphenyl)amino]benzoic acid
MeOH	Methanol
MGC	Mogul Geometry Check
MMP	Matched-Molecular Pairs
N2	Nitrogen
NaH	Sodium hydride
NEt_3	Triethylamine
NMR	Nuclear Magnetic Resonance
Norvir	Commercial name for Ritonavir
NPBA	N-phenylbenzamide
NSAID	Non-steroidal anti-inflammatory drug
olanzapine	2-methyl-4-(4-methylpiperazin-1-yl)-10H-thieno[2,3-b][1,5]benzodiazepine
Pd/C	Palladium on activated charcoal
PIHn	Polymer-Induced Heteronucleation
PIXEL	Lattice energy calculation protocol
PXRD	Powder X-Ray Diffraction
QSAR	Quantitative Structure-Activity Relationship
θ_{thio}	Torsion angle between the two aromatic rings of a ROY-like structure

Ranitidine	1-*N*'-[2-[[5-[(dimethylamino)methyl]furan-2-yl]methylsulfanyl]ethyl]-1-*N*-methyl-2-nitroethene-1,1-diamine
Ritonavir	1,3-thiazol-5-ylmethyl N-[(2S,3S,5S)-3-hydroxy-5-[[(2S)-3-methyl-2-[[methyl-[(2-propan-2-yl-1,3-thiazol-4-yl)methyl]carbamoyl]amino]butanoyl]amino]-1,6-diphenylhexan-2-yl]carbamate
Rochelle's salt	Sodium potassium tartrate
ROY	5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile
ROY-ol	5-hydroxy-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile
RT	Room Temperature
SC-XRD	Single Crystal X-Ray Diffractometry
SFI	Solid Form Informatics
SMILES	Simplified Molecular Input Line Entry System
S_NAr	Nucleophilic aromatic substitution
SPR	Structure-Polymorphism Relationship
TA	Tolfenamic acid
THF	Tetrahydrofuran
TLC	Thin-Layer Chromatography
Z'	Number of molecules contained within the asymmetric unit of a crystal

Table of Contents

Chapter 1 Introduction

1.1 Polymorphism

Polymorphism refers to the ability of a single chemical entity to crystallise into multiple distinct crystalline forms[1]. Two polymorphs of the same compound are chemically identical, but their physical properties can differ significantly due to the different arrangement of molecules within the crystal lattices of the two polymorphs. This variation in physical properties between polymorphs can lead to unexpected behaviour during downstream processes, and can lead to problems in the production of products utilising crystalline materials.

The phenomenon of polymorphism was first documented in organic compounds by Wohler and Leibig in 1832, when multiple distinct crystal habits were observed during an investigation into the crystallisation of benzamide[2]. Throughout the remainder of the 19th century steady progress was made in documenting and understanding the phenomenon of polymorphism, including the first recorded instance of the utilisation of hot-stage microscopy alongside temperature-time curves to observe transitions between the four different polymorphs of ammonium nitrate[3]. This technique would go on to become standard practise in the identification of crystal polymorphs and still finds use today.

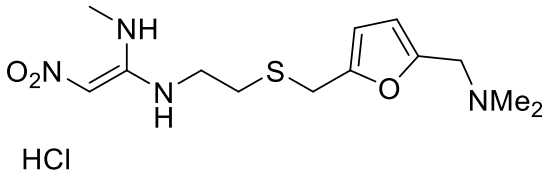
After the initial wave of discoveries related to crystal polymorphism, the field saw a decline in interest. This was largely due to the technical/technological difficulties associated with the characterisation of crystal forms, along with polymorphism being viewed as an interesting quirk of some crystalline systems, rather than more ubiquitous phenomenon. As the technology for rapid crystal structure determination advanced, interest in the field of polymorphism began to see an increase during the second half of the 20th century. In the 1950's, two groups independently published books containing sections discussing the polymorphism of organic materials: the first led by Ludwig Kofler based in the university of Innsbruck[4], and the second led by Walter McCrone, originally based out of Cornell[5].

In 1969 a review was produced by McCrone and Halebian for the Journal of Pharmaceutical Sciences concerning crystal polymorphism and its connection to the pharmaceutical industry[6]. This seminal work detailed many of the nuanced effects arising from the presence of polymorphic forms in a pharmaceutical setting, including the effect of polymorphism on chemical stability[7], bioavailability/generally equivalent dosage forms[8,9] and the tabletting process[10]. Moreover, the review also compiled the many varied methods used to characterise different crystal polymorphs, including X-Ray powder diffraction[11–

[13], differential thermal analysis[14] and the use of phase diagrams[15]. The attention afforded to the review upon release reflected the industrial attitude towards polymorphism at the time. However, a series of high-profile events within the pharmaceutical industry towards the end of the 20th century would see a surge in interest in the field of polymorphism, and this review would become a central text in the understanding of the polymorphism phenomenon.

1.1.1 Industrial Relevance of Polymorphism

There are two main reasons for the industrial relevance of polymorphism which were both brought into the public conscience throughout the 1990s and early 2000s. The first of which is the relevance of polymorphism to the patenting and trademarking of therapeutics, which was highlighted by the patent litigation between Glaxo and Genpharm over Zantac in 1991[16]. Zantac is a histamine-2 blocker used to reduce the production of stomach acid, and was the highest-selling drug in the world in 1991, with sales of $3.45 billion per year. The patent litigation trial focused largely on polymorphism of the active ingredient, and the publicity surrounding the trial and subsequent fallout resulted in an increase in the pharmaceutical relevance of polymorphism, highlighted by the increased number of patents containing the word “polymorphism” after the conclusion of the zantac trial. Ranitidine hydrochloride can exist in two polymorphic forms, “Form 1”, which was under patent until July 1997, and “Form 2”, which was under patent until December 2002. The generics firm Genpharm sought to make a Form 2 product prior to the patent expiration, and also challenged the validity of the Form 1 patent due to ambiguity in the method used to produce crystals. Glaxo and Genpharm reached an agreement in October 1995, in which it is believed that Glaxo likely paid Genpharm $132.5 million[17]. Chemical structure for zantac can be found in Figure 1.



Zantac
(Ranitidine Hydrochloride)

Figure 1; Chemical structure for Zantac (Ranitidine Hydrochloride)

The next and arguably more influential case for the industrial relevance of polymorphism began in 1998 with the release of Norvir, an antiviral medication used in the treatment of acquired immunodeficiency syndrome (AIDS)[18,19]. Towards the end of 1998 it was noted that a number of batches of the slow-release Norvir capsules were failing dissolution tests. This resulted in an

independent investigation into the cause of the failures, during which it was found that the failed batches of Norvir contained a heretofore unseen form of the active ingredient, ritonavir. Norvir was originally formulated using Form I ritonavir, however, analysis of the failed batches revealed that all ritonavir present was in Form II, a more stable and less bioavailable form of ritonavir, which resulted in the failure of the dissolution tests of Norvir. The chemical structure of ritonavir, along with images of Form I and II crystals, can be found in Figure 2.

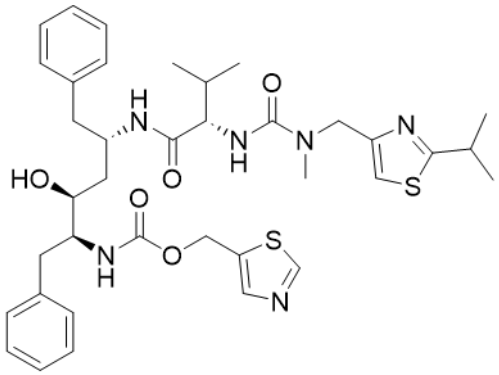


Figure 2; Chemical structure of ritonavir, the active ingredient in the antiviral medication Norvir.

The late-stage discovery of form II ritonavir led to a total product recall and reformulation of Norvir products, a process which was not only expensive, but also cut into the vital post-launch period of Norvir, harming the overall momentum of the product.

The Norvir/ritonavir example served as a warning of the dangers of late-stage polymorph discovery on the performance of pharmaceutical products, and led to the proliferation of techniques used to explore the polymorphic landscape of active ingredients in order to avoid cases such as ritonavir. This led to the situation observed across the pharmaceutical industry today, where billions of dollars and countless working hours are spent utilising a range of experimental and computational techniques in order to identify and isolate the optimal form of active ingredients before they are formulated into therapeutic products.

1.1.2 Stability Relationships Between Polymorphs

Arguably the most important physicochemical property of a polymorphic system, particularly in a pharmaceutical setting, is the thermodynamic stability relationship between the different polymorphic forms. Many physicochemical properties, such as solubility and bioavailability, are inherently tied to the thermodynamic stability of a given crystal form. Moreover, formulations which contain crystalline material are engineered specifically to the properties of the crystal form present, and as such, any post-production change in the crystalline

phase present can significantly reduce the efficacy of the formulated product (i.e. the ritonavir case).

When endeavouring to formulate a product containing a crystalline active ingredient it is preferred that the crystal is present in its most thermodynamically stable form (i.e. the form with the lowest Gibbs free energy), in order to mitigate the risk of storage- or process-mediate phase transitions. It is therefore important to establish the stability relationship for the polymorphic system at this stage.

Stability relationships for polymorphic forms fall into two categories: monotropic polymorphs, whose stability relationship is independent of temperature, and enantiotropic polymorphs, whose stability relationship is temperature-dependent. From control is generally simple in the case of monotropic polymorphic systems, however, enantiotropic systems present a challenge, as the most stable form at ambient temperatures (i.e. normal storage conditions) may not be the most stable at all temperatures.

Enantiotropic polymorphs can reversibly interconvert between crystalline phases without entering the liquid or gaseous states, and are characterised by a transition temperature (Tt), above and below which the stability order of the two polymorphs is reversed. It is important to study the stability relationships between polymorphic forms over a range of temperatures, and in the case of enantiotropic polymorphs, to establish the value of Tt and the influence this has on relative polymorphic stability.

1.1.3 Polymorph Screening

Since the turn of the millennium polymorph screening has become an integral step in the development of any product utilising crystalline materials as their active ingredient. Polymorph screening is a broad and diverse field utilising a range of different techniques in order to explore the crystallographic landscape of an active ingredient, and the end-goals of a given screen dictate the methods used and experiments performed.

In a pharmaceutical setting, two of the primary concerns are bioavailability and shelf-life of the active ingredient upon formulation into a product, as highlighted by the Norvir/ritonavir case[18]. As such, the focus of many polymorph screens performed in the pharmaceutical industry is on isolating the most thermodynamically stable form(s) of the active species[20]. The use of the most thermodynamically stable form of a compound avoids the issue presented by ritonavir, in which the late-stage discovery of the more thermodynamically stable form II ritonavir led to a complete product recall and reformulation.

The methods employed in a polymorph screen vary depending on which aspects of the polymorphic landscape are being explored. Typical polymorph screens utilise crystallisations from melts, vapours along with the evaporation of solvents at different temperatures and/or the addition of antisolvents[21]. Other methods utilised include slurrying of seed materials at different temperatures in order to isolate stable forms[22]. More exhaustive methods used to explore as wide a range of the polymorphic landscape as possible include the use of polymer-induced heteronucleation (PIHn), in which hundreds of unique polymers are used to in order to grant kinetic access to as many polymorphs as possible[23], whilst other methods utilise microencapsulation in order to access high-energy, metastable polymorphs[24,25].

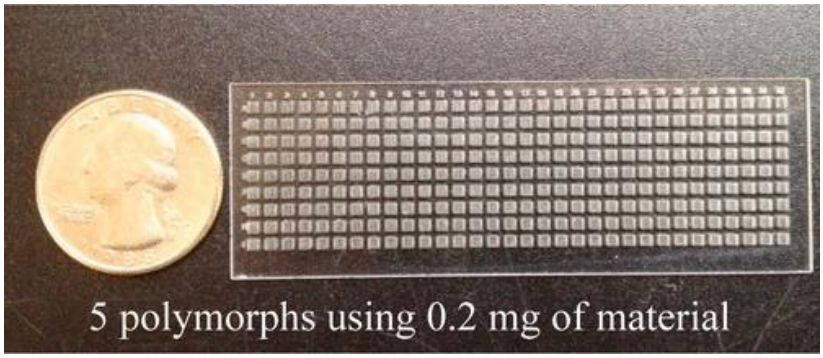


Figure 3; the polymer-loaded plate used in PIHn polymorph screening, taken from [23] with permission from the rights holder.

Table 1 contains a list of techniques commonly employed to study solid-state properties. Detection of polymorphs during a polymorph screen can be performed using many different methods used to characterise solids, however the specific techniques employed in a given polymorph screen depend on the key parameters which are to be determined. Some characteristic parameters include solubility, thermal events, crystallographic properties and chemical environment. It is typically advised that two or more complimentary characterisation techniques are employed during polymorph detection in order to obtain reliable knowledge of the forms present. However, due to the large number of samples produced *via* modern screening methods, the use of multiple characterisation steps may be unreasonable due to the time this would require. As such, modern high-throughput screens may prefer to employ hyphenated techniques for analysis, which would allow complementary form data to be collected in a single step.

Table 1; Methods used to study solid-state properties, commonly employed as detection methods in polymorph screening

Technique	Data Measured	Property Characterised
Differential Scanning Calorimetry (DSC)	Heat flow vs temperature	Thermal Events
Infrared (IR) Spectroscopy	IR spectrum	Chemical information
Raman Spectroscopy	Raman spectrum	Chemical information
Solid State Nuclear Magnetic Resonance (NMR) Spectroscopy	Magnetic resonance	Chemical information
Solubility/Dissolution	Amount of material dissolved in different solvents/temperature vs time	Measurement of solubility/dissolution rate
Terahertz Pulsed Spectroscopy (TPS)	Terahertz Pulsed Spectrum	Chemical information
Thermogravimetric Analysis (TGA)	Change of mass vs temperature	Solvate/hydrate presence
X-Ray Diffraction (XRD) (Single Crystal and Powder XRD)	Diffractogram	Crystallographic properties

1.1.4 Crystal Structure Prediction

The ubiquity of polymorphism and hence the necessity of polymorph screening leads to significant time and effort being directed towards polymorph screening in industry. As such, there has been a widespread interest in the development of tools which could supplement or guide polymorph screens utilising the powers of modern computational power. One such approach which shows promise is known as crystal structure prediction (CSP).

CSP is a computational approach used to produce a small, yet complete, set of crystal structures which are likely to be accessible experimentally[26]. CSP is a powerful technique used to compliment experimental screening in both the agrochemical and pharmaceutical industries[27], allowing for the interpretation of experimental results[28], along with assisting with the resolution of structures from powder X-ray diffraction (PXRD)[29] or nuclear magnetic resonance (NMR)[30] data.

CSP is used to generate a set of candidate structures with the goal of exploring the lattice-energy landscape, which is used to search for the global energy minima. However, the computational cost of calculating all structures to the accuracy required for accurate predictions would be too great, and a multi-stage approach of increasing model complexity is used on a sequentially refined set of candidate structures. Lattice energy thresholds (typically ΔH_{lat} at 0 K, 0 Pa) are used to refine the structure set and filter out high energy structures. Post-refinement, the remaining candidate structures are recalculated using a lattice energy model of higher complexity and the global energy minima are located. It is typical to perform this refinement one or two times, before a final assessment is then applied to the structures remaining after the final refinement stage using a highly accurate model of the lattice or free energy.

Models for CSP usually utilise two primary methods for structure and lattice energy calculation. The first utilises force field models and are typically employed during the early stages of CSP studies, due to their relative computational simplicity when compared to other methods, such as electronic structure methods. Force field methods split the lattice energy of a crystal, ΔU_{lat}, into intramolecular components and intermolecular components, with the force fields employed being tailored to the most prevalent physical properties of the system.

$$\Delta U_{lat} = \Delta E_{Intra} + E_{Inter}$$

Electronic structure approaches allow for the consideration of molecular interactions throughout the whole crystal, with the lattice energy of the crystal being calculated from atomic positions from within the crystal lattice. Periodic density functional theory (DFT) methods are typically preferred to other quantum mechanical methods due to being seen as a good compromise between computational cost and predictive accuracy[26].

CSP has shown promising developments over the past decade, highlighted by the sixth blind CSP challenge, hosted by the Cambridge Crystallographic Data Centre (CCDC) in 2016[31]. This challenge consisted of five target systems: a small rigid molecule, a polymorphic former drug candidate, a chloride salt hydrate, a cocrystal and a bulky flexible molecule. A wide number of participants utilising a range of different CSP methodologies participated in the challenge, resulting in the prediction of all but one of the target structures, a disordered structure for which Z' = 2. The challenge demonstrated the growing maturity of the field of CSP, with progress made in the determination of flexible molecules, the establishment of best-practises for CSP and the usage of hierarchical approaches to structure ranking particularly highlighted.

However, a number of key issues regarding CSP remain. Particularly, the difficulty in generating structures for molecules with a high degree of molecular flexibility was noted. Furthermore, the issue of producing definitive energetic ranking of structures proves challenging. Ultimately, CSP is a powerful technique which can serve as an invaluable resource in the exploration of the polymorphic landscape of a given compound. However, the computational resources required, along with the inability to provide a definitive structure ranking provide a barrier to the widespread/general application of the technique.

1.2 Solid-Form Informatics

In the context of drug discovery, computational methods such as virtual screening, molecular docking and quantitative structure-activity relationship (QSAR) models have greatly improved the efficiency of the drug discovery process[32,33]. This is achieved by reducing the cognitive and experimental load on the researchers and placing it on computational and knowledge-based models. Similar computational approaches exist to supplement the drug development process, which look to reduce the experimental workload by exploring the solid-form landscape of compounds *in-silico.*

Drug development and solid-form selection are experimentally taxing and expensive processes, as care must be taken to ensure the solid-form landscape is thoroughly understood before proceeding with the nominated form for further development. This is due to the late-stage appearance of stable polymorphs being highly disruptive to production, such as was the case with Norvir (ritonavir). This led to an expensive product recall and reformulation due to the appearance of undiscovered, highly-stable forms post-production[34]. Moreover, polymorph screen data are required when submitting dossiers for drugs awaiting FDA approval.

Computational methods have the ability to streamline/optimise the solid-form selection process by providing insight into the solid-form landscape of a given compound before exploring the landscape experimentally. *Ab-initio* methods such a CSP are powerful techniques which can tell researchers about the relative stability of theoretically accessible crystal structures, helping researchers design experimental screens and allowing them to target lower-energy forms and hence select the optimal form to take forward into production[35]. However, ab-initio methods are computationally expensive and time consuming, making them less suitable for routine use earlier in drug development processes.

An alternative computational approach to aid drug development and solid form selection is to utilise databases of structural information, such as the CSD[36], in

order to produce knowledge-based models to assess a compound's propensity towards polymorphism. These methods utilise the vast data stored in these databases along with statistical modelling to make predictions about the stability and polymorphic propensity of structures, based on patterns found within the database. Such approaches, also known as informatics-based approaches, have the power to provide insight into a given solid form of a compound, and help to assess the likelihood of alternate forms being accessible. This information can then be used to design an effective screening programme to discover alternative solid-forms and reduce the time and money spent during this process.

One such approach is known as Solid-Form Informatics (SFI). Introduced by Bob Doherty in 2004 to describe the process of using structural knowledge to inform key decisions in pharmaceutical development[37], the concept was expanded upon by Galek et al. in 2011 to encapsulate a knowledge-based approach to evaluate crystal structures and predict crystal properties[38]. SFI utilises the CSD and its 1.25 million crystal structures to create knowledge-based models in order to assess the intramolecular, intermolecular and supramolecular environment of a crystal.

Through the use of SFI, Galek et al. sought to answer the questions:

"Is my crystal structure reliable?"

"Is my crystal likely to crystallise in a needle- or plate-like morphology?"

"How likely are alternative polymorphs?"

For a given input structure, the ultimate goal of SFI is the production of a fully-automated "Structure Report". This report would rapidly and automatically describe, analyse and characterise a given crystal structure, highlighting any structural or chemical issues.

1.2.1 Case Study: lamotrigine

In order to demonstrate the capabilities of SFI, Galek et al. selected the drug compound lamotrigine, which had recently been uploaded as the 500000th entry to the CSD. It was first synthesised at Wellcome laboratories as a potential epilepsy treatment, where it was expected to behave as a dihydrofolate reductase inhibitor. Though it was found to display little antifolate activity, it was found to display significant anticonvulsant activity, and was approved by the American Food and Drugs Administration (FDA) for use in this area in 1994 [39]. It was also found to play a role in the management of mood disorders, and was subsequently approved for the treatment of bipolar disorder in 2003[40,41].

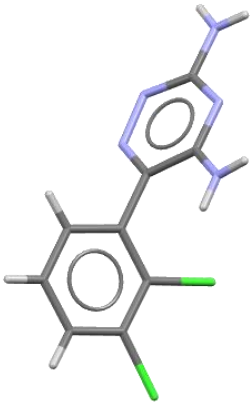


Figure 4; chemical structure for a molecule of lamotrigine as found in EFEMUX01. Image was generated using Mercury 2023.2.0 (Build 382240), CSD v5.41.

The crystal structure of lamotrigine was first published in 2009 by Sridhar and Ravikumar[42], before being uploaded to the CSD. It is a relatively typical drug molecule, with its prominent physicochemical properties (Log P, number of H-bond donors/acceptors) adhering to Lipinski's rule of five[43].

Galek et al. used lamotrigine as an example of how the knowledge contained within the CSD can assist in understanding the properties of the solid forms of compounds. In this work, the SFI analysis performed by Galek et al. is repeated using the most up to date version of the CSD and its modules. Each step in the SFI analysis is evaluated for its ability to answer the three questions stated previously, along with its ability to assess polymorphic landscape of lamotrigine *in-silico.*

1.2.1.1 Intramolecular Environment

Intramolecular geometry is one of the many factors involved in the formation of a crystal structure. Within a crystal molecules assemble into regular packed arrangements which repeat infinitely along each of the cell axes. Molecular confirmation and intramolecular geometry dictate how well each molecule can pack within the crystal structure, along with influencing which intermolecular interactions can form. Moreover, irregularities in intramolecular geometry can be an indication of issues during structural determination, and that a redetermination may be necessary.

Compounds are stored within the CSD as CIFs (crystal information files), allowing for the representation of lamotrigine in many forms (2D structure, SMILES string, InChI etc.) via the use of the program *Decifer.* This allows for rapid and effective mining of the CSD for compounds containing the same or similar motifs. A

question in the SFI process introduced by Galek et al. is whether the 3D structure is represented accurately in the given CIF. To probe this, they used the program *Mogul*[44], found within the CSD-Core tab of the *Mercury*[45] interface.

Mogul Geometry Check is an informatics tool found within the Mogul module in Mercury, which is used to compare selected structural descriptors for an input structure against compounds containing the same structural motif within the CSD. These structural descriptors are Bond Length, Valence Angle and Torsion Angle, and allow for rapid assessment of intramolecular geometry based on how well these values follow the trends observed across the CSD.

Table 2; output for the Mogul Geometry Check performed on EFEMUX01 using CSD v5.41

Bond	17 bonds, 0 unusual
Angle	24 angles, 0 unusual
Torsion	4 flexible torsion angles, 0 unusual
Ring Geometry	2 rings, 0 unusual

Performing the Mogul Geometry Check on the pure crystal form of lamotrigine (EFEMUX01) provides us with an instant picture of the bond lengths, angles and torsions and whether they follow the trends of the CSD. Table 1 summarises these data and shows that the bond angles, lengths and torsions found within EFEMUX01 all follow the established trends observed in the CSD. Mogul Geometry Check does this by breaking the input structure (EFEMUX01) into chemical fragments and calculating the geometric descriptors for these fragments based on the 3D information contained within the CIF. The CSD is mined for examples of this fragment and the geometric descriptor is calculated for each. Next, a histogram plot is produced for these fragments and their associated geometric descriptors which displays how these descriptors vary across the CSD, with green bins and their relative height describing the most likely values for these descriptors and their statistical likelihood. Finally, the value for the descriptor in the input structure is indicated on the histogram plot as a red line. If this line falls within one of the bins of the histogram the descriptor is labelled as *not unusual*, meaning there are sufficient examples of this value observed within the CSD. If the line falls outside of the bins of the histogram, the descriptor is labelled as *unusual*. *Unusual* descriptor values could indicate that a structural redetermination is required, or alternatively that the molecules within the structure sacrifice a favourable intramolecular geometry due to packing.

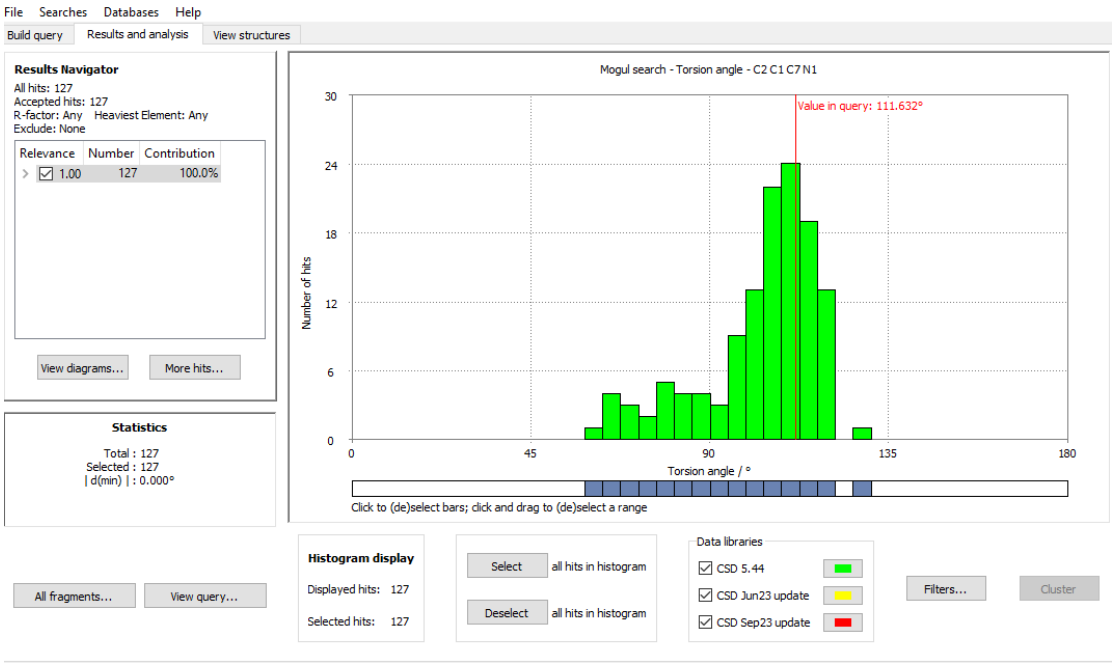


Figure 5; histogram displaying the statistics for the distribution of torsions about the C2-C1-C7-N1 fragment in the CSD v5.41. Red line indicates the value of the C2-C1-C7-N1 torsion in the input molecule EFEMUX01. Image generated using Mercury 2023.2.0 (Build 382240), CSD v5.41, using the Mogul Geometry Check feature located within the "CSD-Core" tab.

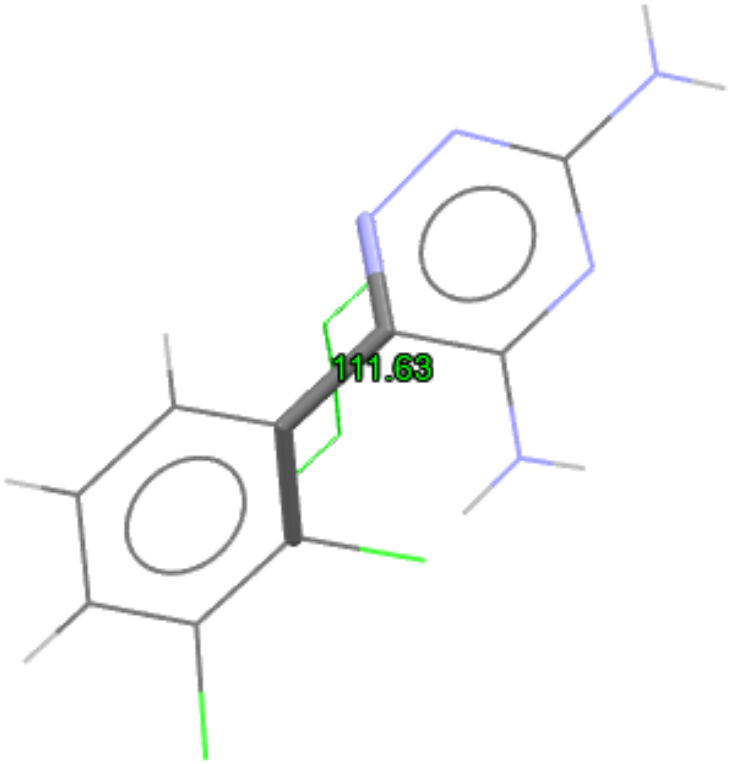


Figure 6; visualisation of the C2-C1-C7-N1 torsion in a lamotrigine molecule in EFEMUX01. The value of the torsion angle (111.63º) along with the angle in question are highlighted in green. Image generated using Mercury 2023.2.0 (Build 382240), CSD v5.41, using the Mogul Geometry Check feature located within the "CSD-Core" tab.

Analysis of the structure and descriptors for geometrical properties of lamotrigine serves as the first step of understanding the crystalline environment it exists within. Through statistical analysis of the intramolecular geometry we can be confident that the structural representation of EFEMUX01 is correct[46], and that there are no unusual intramolecular geometries due to the presence of unusual chemistry[47].

1.2.1.2 Intermolecular Environment

Expanding our analysis beyond the individual molecule within the crystal, it is useful to begin with an overview of the crystal structure as a whole. Table 2 summarises the basic structural information associated with the EFEMUX01 determination of lamotrigine. There is nothing unusual with respect to the unit cell parameters of the structure, and the low R-factor suggests that the refined model provided a good fit to the experimentally acquired data. The packing coefficient associated with the structure (0.72) is typical for an organic molecule (around 0.71)[48], and no void space was found within the structure, hence there are no clear suggestions of instabilities within the crystal structure, such as missing solvent atoms.

Table 3; structural information for the EFEMUX01 form of lamotrigine.

Unit Cell Details	Values	Structure Details	Values
a / Å	19.136(3)	R-factor / %	2.81
b / Å	8.6409(12)	Density / g cm-3	1.607
c / Å	13.5549(18)	Temperature / K	294
a / º	90	Colour	Colourless
b / º	109.172(2)	Habit	Block
g / º	90	Packing coefficient	0.72
Cell Volume / Å^3	2117.02	Void Space / %	0.0
Space Group	C 2/c		
Z Values	Z: 8, Z': 1		

The next step Galek et al. perform is to investigate the non-covalent interactions occurring between neighbouring molecules within the crystal structure. This is done to evaluate the local environment of the molecule and determine whether it is energetically favourable, and if any packing inefficiencies can be rationalised through understanding of these non-covalent interactions. Galek et al. use the *IsoStar*[49] functionality within Mercury to perform this analysis. IsoStar is a collection of databases detailing the distance and angle data between a central group (some functional group i.e. the dichlorobenzene ring in lamotrigine) and a selection of probe groups from within the CSD. This mines the CSD for all intermolecular contacts involving a selected functional group present in the input molecule, in this case lamotrigine (EFEMUX01). These produce colour-coded scatterplots detailing the statistical likelihood that a selected probe group (i.e. uncharged NH hydrogen, aromatic Cl group etc.) will be found at a given position around the central group. The local density of points for a certain group at a certain position gives a measure of propensity towards forming an interaction with that group at this position. Thus, we can infer that interactions formed at these points are energetically favourable.

However, in the years since publication many advancements to the CSD have been made which sought to improve the ease and efficacy of assessing the non-covalent interactions within crystal structures. One of these advancements was a direct evolution of the IsoStar concept, known as *Full-Interaction Maps*[50] (FIMs).

FIMs combine the IsoStar libraries and uses them to assess an input molecule as a whole. By selecting probe molecules, maps of hotspots can be produced around the molecule detailing the statistical likelihood of finding a given probe at a certain point. The maps are built around the molecule taking into consideration intramolecular geometry and steric effects. The FIMs for lamotrigine using uncharged NH hydrogen, carbonyl oxygen and aromatic CH hydrogen are provided in Figure 4. FIMs are an advancement of IsoStar plots as both are built on the same data, but FIMs can be used to consider the molecule as a whole (rather than fragments of functional groups), and multiple probe groups can be mapped at the same time. This allows for the competition between two functional groups within the structure to be more easily understood.

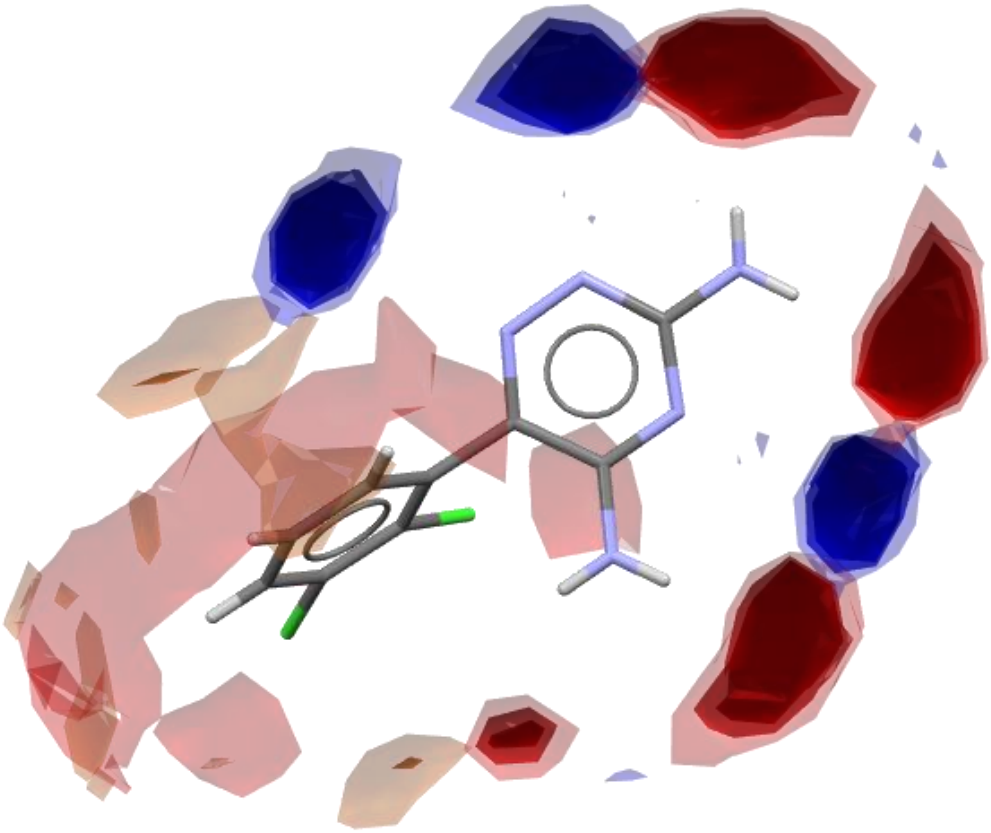


Figure 7; Full Interaction Maps (FIMS) generated or the lamotrigine molecule in EFEMUX01 using uncharged NH hydrogen (red) carbonyl CO oxygen (blue) and aromatic CH hydrogen (brown) as the probe groups. Shade of hotspot determines statistical likelihood of finding probe at a given position, with darker spots presenting a higher statistical chance. Image generated using Mercury 2023.2.0 (Build 382240), CSD v5.41, using the Full Interaction Maps feature located within the "CSD-Materials" tab.

In order to assess the non-covalent interactions formed by the lamotrigine molecules whilst packing in the EFEMUX01 and evaluate their favourability, Mercury can be used to expand the closest interatomic distances between our central molecule and its nearest neighbours within the crystal. Coupled with the

FIMs, this feature can be used to assess the intermolecular interactions based on: i) which atoms/groups are involved in the interaction and ii) what position (distance and angle) is the interaction situated at.

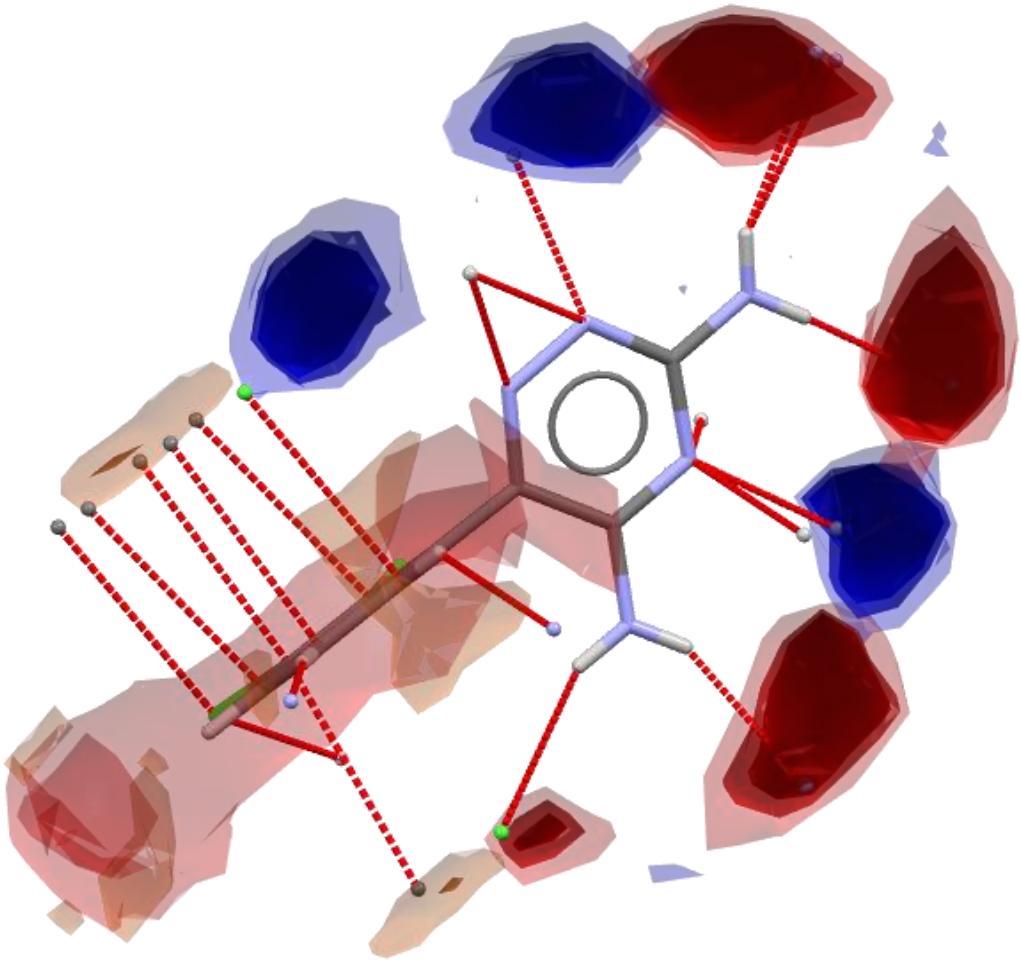


Figure 8; intermolecular short contacts between the central lamotrigine molecule in EFEMUX01 and its nearest neighbours from within the crystal structure, overlaid with the calculated FIMs hotspots. Image generated using Mercury 2023.2.0 (Build 382240), CSD v5.41, using the Full Interaction Maps feature located within the "CSD-Materials" tab.

The first thing to note when looking at the intermolecular interactions formed by molecules within the EFEMUX01 form of lamotrigine is that they mostly fall well within the hotspots predicted by FIMs. This suggests that these are interactions with a high propensity of formation, and hence we can infer they are energetically favourable. The intermolecular interactions formed by molecules in EFEMUX01 are primarily hydrogen-bonds, with a significant π-π stacking interaction observed above the plane of the dichlorobenzene moiety. One way to measure the strength and stability of a hydrogen-bond is to measure its length and contact angle. Hydrogen-bonds have well-defined bond angle criteria, with X-H --- Y bonds

usually found at 180°, and deviation from this criteria is typically associated with a weaker hydrogen-bond[51]. The distance and angle data for the intermolecular hydrogen-bonds formed within the EFEMUX01 crystal are provided in Table 3.

Table 4; intermolecular hydrogen-bond donor atoms, acceptor atoms, distance and angle statistics for the central lamotrigine molecule in EFEMUX01.

Atom X	Atom Y	X---Y Distance / Å³	X---Y Angle / °
N3 – H5	N4	2.330	176.78
N5 – H6	N2	2.399	151.02
N3 – H4	N1	2.539	155.04
N5 – H7	Cl2	2.648	150.95
N3 – H4	N2	2.737	131.47

The molecules in EFEMUX01 form multiple intermolecular hydrogen bonds, with the N3-H5 --- N4 bond being the strongest. This bond forms a homodimer with a rotated molecule of lamotrigine which helps to satisfy the donor-acceptor-donor hotspots present in the FIMs. The N3 – H5 --- N4 hydrogen-bond dimer can be found in Figure 6.

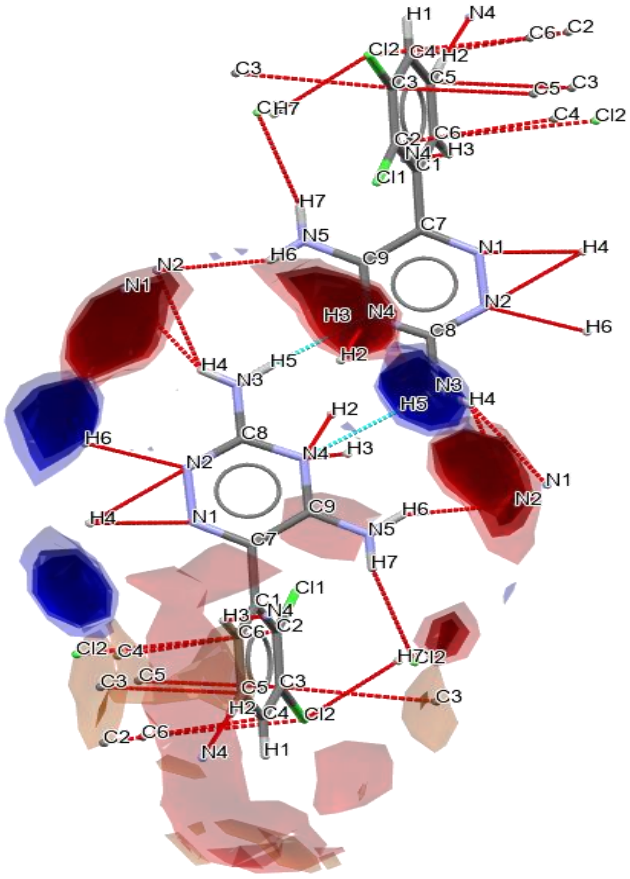


Figure 9; visualisation of the strongest intermolecular hydrogen-bond formed by lamotrigine in EFEMUX01 (N3-H5---N4) labelled in cyan. All other intermolecular short contacts formed by lamotrigine in EFEMUX01 are displayed in red. FIMs are overlaid to show the most likely interaction of N4 is satisfied. Image generated using Mercury 2023.2.0 (Build 382240), CSD v5.41, using the Full Interaction Maps feature located within the "CSD-Materials" tab.

The other significant non-covalent interaction formed by molecules in EFEMUX01 is an attractive π-π stacking interaction between dichlorobenzene motifs of neighbouring molecules. These interactions are also found in the same position as the aromatic CH hotspot (brown) generated via the FIMs. This suggest that this interaction has a high propensity of forming, and hence has a stabilising effect on the structure.

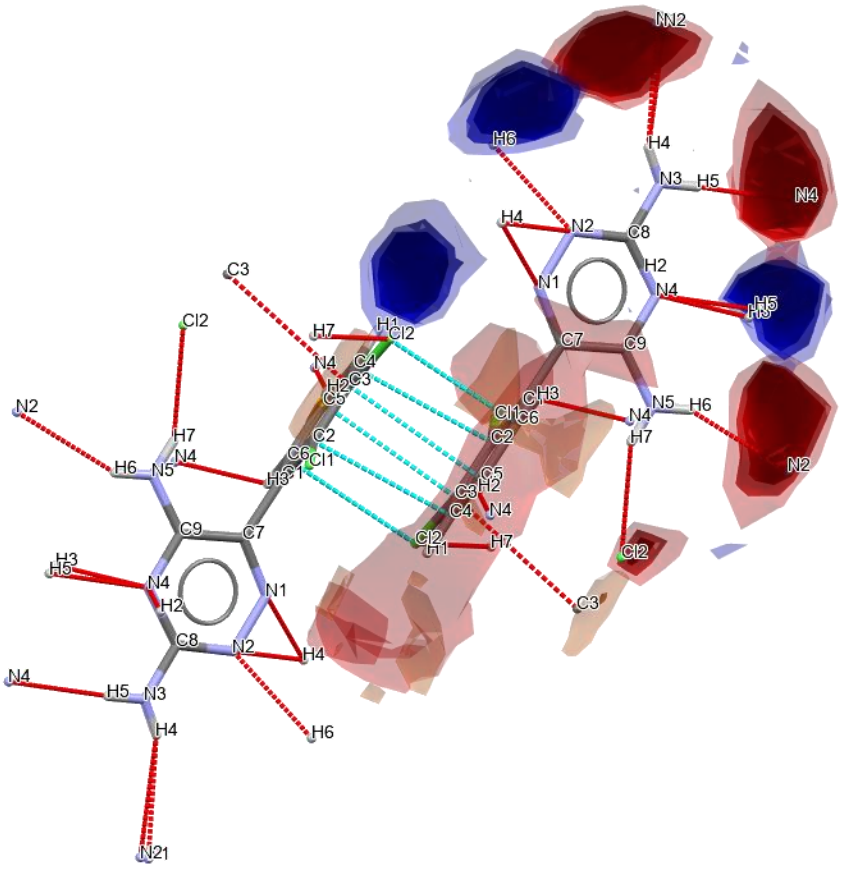


Figure 10; visualisation of the π-π stacking interaction observed between neighbouring lamotrigine molecules in EFEMUX01 highlighted in cyan. FIMs overlaid to demonstrate the aromatic CH hydrogen hotspot (brown) is satisfied by this interaction. Image generated using Mercury 2023.2.0 (Build 382240), CSD v5.41, using the Full Interaction Maps feature located within the "CSD-Materials" tab.

Analysis of the intermolecular interactions formed by molecules in EFEMUX01 finds that the structure is likely to form many hydrogen bonds, due to the good balance and high number of hydrogen-bond donors and acceptors in the molecule. Comparison of the FIMs to the observed hydrogen-bonds within the structure displays that all strong donors and acceptors in the molecule are satisfied. Moreover, most hydrogen-bond donors and acceptors in the molecule have their interactions satisfied by interactions with a high statistical likelihood of formation. The only exception being the donor N1, which has its interaction satisfied by a bond with N3 – H4, which is also sharing electron density with N2 in a significantly weaker hydrogen-bond . This interaction is found outside of the hotspot predicted by the FIMs, leaving this hotspot unsatisfied. The other non-covalent interaction observed between molecules within EFEMUX01 is a π-π stacking interaction above and below the plane of the dichlorobenzene moiety.

These interactions also fall within the hotspots of the FIMs, and hence are thought of as favourable interactions.

This analysis suggests that the intermolecular interactions within EFEMUX01 are satisfied, and that the crystal forms is likely to be stable. Analysis of the interactions in the context of the trends observed in the CSD suggest a high degree of satisfaction. However, one hydrogen bond acceptor group forms a hydrogen-bond which is significantly different to trends within the CSD, and hence could be satisfied by a more favourable hydrogen-bond which exists within the hotspot predicted by FIMs. This could suggest that alternative polymorphic forms of lamotrigine could exist with this more-favourable hydrogen-bond.

1.2.1.3 Supramolecular Environment

By viewing the lamotrigine molecule in EFEMUX01 we have been able to describe its intramolecular structure, and by exploring the non-covalent interactions formed by each molecule within the crystal we have been able to describe the intermolecular structure of molecules in EFEMUX01 lamotrigine crystals. However, the physicochemical properties of the bulk crystal are dependent on how these intermolecular interactions accumulate across the whole of the crystal and its faces to form the internal network, or supramolecular structure, of the crystal.

As we have discussed in section 1.2.1.2, lamotrigine molecules in EFEMUX01 form several intermolecular hydrogen-bonds to neighbouring molecules within the crystal. A useful tool contained within the CSD for assessing hydrogen-bonding within crystal structures is the *Hydrogen Bond Propensity*[52] (HBP) module. HBP utilises logistical regression along with the hydrogen-bonding data present in the CSD to evaluate the coordination and propensity of each donor and acceptor atom in a given structure. It does this by first inputting the structure of a compound to model (in our case EFEMUX01) and identifying the feasible donor and acceptor atoms. A competition function then evaluates the balance of available lone pairs and hydrogen atoms for donation, and a steric density function evaluates the steric bulk around each donor and acceptor atom. These inputs are then compared to a bespoke dataset derived from the CSD which contains the relevant donor and acceptor atoms (and hence the ability to produce similar hydrogen bonding networks), and a statistical model of the potential hydrogen-bonding networks possible for the donors and acceptors present is constructed. These potential networks are ranked based on their coordination

score and propensity, with higher average coordination and propensity scores suggesting more efficient and satisfied hydrogen-bonding networks overall.

The input structure and hydrogen-bond donor and acceptor fragments can be found in Figure 8.

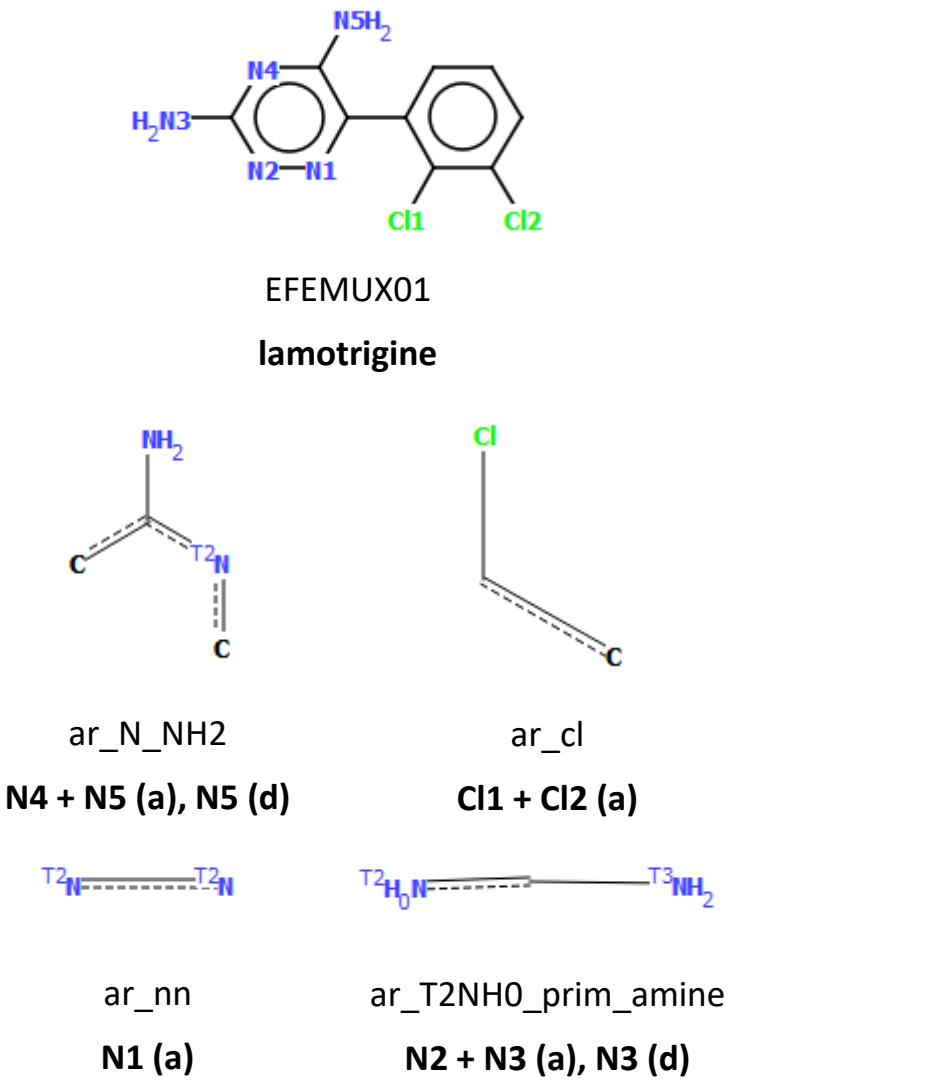


Figure 11; Top) 2D structural representation of lamotrigine used as input for the HBP model. Nitrogen atoms highlighted in blue, chlorine atoms highlighted in green. (Bottom) 2D Representations of lamotrigine functional groups used to construct the HBP model. Names of these functional group representations are given in bold. Below, atom and hydrogen-bond donor (d) or acceptor (a) labels are provided. Images generated using Mercury 2023.2.0 (Build 382240), CSD v5.41, using the Hydrogen Bond Propensities feature located within the "Polymorph Assessment" section of the "CSD-Materials" tab.

There is a reasonable balance between the number of acceptor groups (five) and the number of available hydrogen-bonds (four), suggesting EFEMUX01 is capable of forming complex, multidimensional hydrogen-bonding networks. These donor and acceptor fragments, and the hydrogen-bonding network they form, are then compared to a bespoke dataset derived from the CSD of molecules containing the same functionality. For this dataset each donor and acceptor atom is evaluated and a *coordination score* and *propensity score* is calculated for each donor and acceptor pair.

The coordination score is calculated for each hydrogen-bond donor and acceptor atom within the input structure, and provides a score between 0.00 and 1.00 representing the statistical likelihood that the given donor/acceptor atom will form zero, one or two intermolecular hydrogen-bonds. These scores are calculated based on the regression model produced on the bespoke dataset used to train the model, with a score of 1.00 representing a degree of coordination perfectly in line with the trends of the bespoke dataset, and a score of 0.00 representing a coordination that was not representative of the dataset used to train the model. As we infer that the most commonly observed hydrogen-bonds are the most favourable/strongest, a higher coordination score represents a move favourable hydrogen-bonding network.

Co-ordination scores

(To refresh table: left-click chart point)

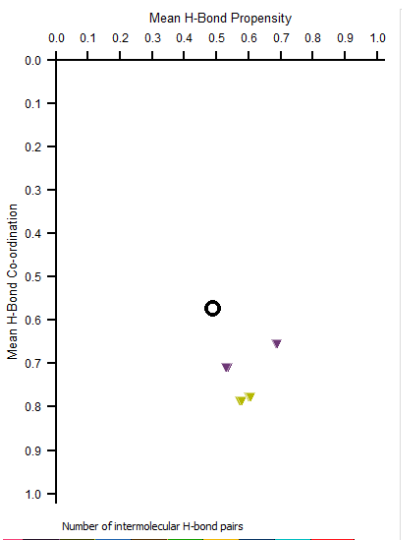
	Atom (D/A)	= 0	= 1	= 2	= 3
1	N3 of ar_T2NH0_prim_amine (d)	0.016	0.241	0.695	**0.048**
2	N5 of ar_T2NH0_prim_amine (d)	0.046	0.275	**0.654**	0.025
3	Cl1 of Cl (a)	**0.956**	0.043	0.001	0.000
4	Cl2 of Cl (a)	0.948	**0.047**	0.004	0.000
5	N1 of ar_nn (a)	0.365	**0.587**	0.048	0.000
6	N2 of ar_T2NH0_prim_amine (a)	0.069	0.844	**0.087**	0.000
7	N3 of ar_T2NH0_prim_amine (a)	**0.948**	0.052	0.000	0.000
8	N4 of ar_T2NH0_prim_amine (a)	0.112	**0.840**	0.049	0.000
9	N5 of ar_T2NH0_prim_amine (a)	**0.989**	0.011	0.000	0.000

ture(s): EFEMUX01 group - 1 item

Figure 12; table of coordination scores for each hydrogen-bond donor and acceptor atom in EFEMUX01 lamotrigine as calculated by the HBP model. The coordination scores for the hydrogen-bonding network present in EFEMUX01 are highlighted on the table. Green scores represent hydrogen-bond environments which follow the trends of the CSD, red scores represent environments which do not follow trends of the CSD. mages generated using Mercury 2023.2.0 (Build 382240), CSD v5.41, using the Hydrogen Bond Propensities feature located within the "Polymorph Assessment" section of the "CSD-Materials" tab.

The hydrogen-bonding network in EFEMUX01 contains five intermolecular hydrogen bonds, and the average coordination score for the hydrogen-bonding network is 0.489, which is a reasonably well satisfied network based on the trends of the CSD. This network does include multiple bifurcated hydrogen-bond donors and acceptors, which are less common within the CSD. The N3 donor atom of ar_T2NH0_prim_amine is involved in three intermolecular hydrogen-bonds, producing a coordination score of 0.048. This donor is in a particularly unfavourable state, which could be seen as a driving force for accessing alternate polymorphic forms with more favourable hydrogen-bonding networks. The HBP

model predicts several hydrogen bonding networks which produce a higher coordination score than the one present in EFEMUX01.



Propensity scores

Hydrogen-bond propensities for EFEMUX01 calculated using 'logit_model_1'

Intermolecular | Intramolecular

	Donor	Acceptor	Propensity	H-Bonds	Donor ranking	Acceptor ranking
1	N5 of ar_N_NH2	N1 of ar_nn	**0.74**	0	1	1
2	N5 of ar_N_NH2	N2 of ar_T2NH0_prim_amine	**0.71**	1	1	2
3	N3 of ar_T2NH0_prim_amine	N1 of ar_nn	**0.66**	1	2	1
4	N3 of ar_T2NH0_prim_amine	N2 of ar_T2NH0_prim_amine	**0.63**	1	2	2
5	N5 of ar_N_NH2	N4 of ar_N_NH2	**0.44**	0	1	3
6	N3 of ar_T2NH0_prim_amine	N4 of ar_N_NH2	**0.35**	1	2	3
7	N5 of ar_N_NH2	Cl2 of ar_cl	**0.09**	1	1	4
8	N5 of ar_N_NH2	Cl1 of ar_cl	**0.07**	0	1	5
9	N3 of ar_T2NH0_prim_amine	Cl2 of ar_cl	**0.06**	0	2	4
10	N3 of ar_T2NH0_prim_amine	Cl1 of ar_cl	**0.05**	0	2	5
11	N5 of ar_N_NH2	N3 of ar_T2NH0_prim_amine	**0.02**	0	1	6
12	N5 of ar_N_NH2	N5 of ar_N_NH2	**0.01**	0	1	7
13	N3 of ar_T2NH0_prim_amine	N3 of ar_T2NH0_prim_amine	**0.01**	0	2	6
14	N3 of ar_T2NH0_prim_amine	N5 of ar_N_NH2	**0.01**	0	2	7

Figure 13; (Left) Plot of average coordination score vs average propensity score for the possible hydrogen-bonding networks of EFEMUX01 lamotrigine as calculated by HBP. White cercle represents network present in EFEMUX01. Purple triangles represent hydrogen-bonding networks with three intermolecular hydrogen-bonds. Yellow triangles represent hydrogen-bonding networks with two intermolecular hydrogen-bonds. (Right) Table of propensity scores calculated for each donor/acceptor pairing in lamotrigine. The donor/acceptor pairs observed in EFEMUX01 are highlighted blue. mages generated using Mercury 2023.2.0 (Build 382240), CSD v5.41, using the Hydrogen Bond Propensities feature located within the "Polymorph Assessment" section of the "CSD-Materials" tab.

The network with the highest coordination score forms three intermolecular hydrogen-bonds and produces an average coordination score of 0.787. This network results in donor/acceptor atoms adopting environments much more in line with the trends of the CSD, with the lowest coordination score being for the N5 donor atom of ar_T2NH0_prim_amine being 0.275. This suggests that

lamotrigine could crystallise in alternate forms to EFEMUX01, in order to form a more stabilising hydrogen bonding network.

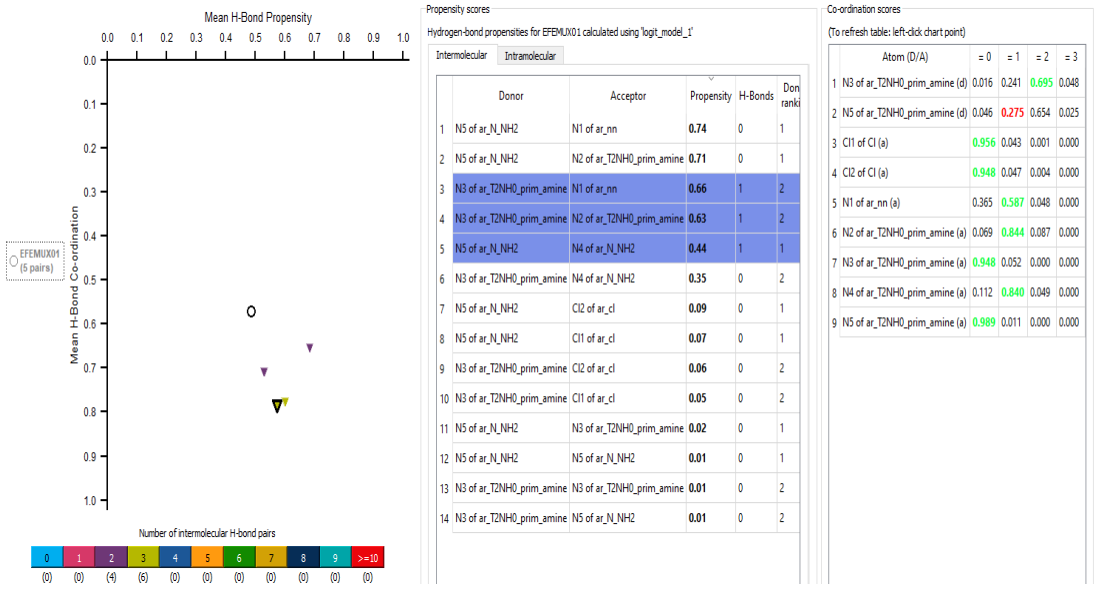
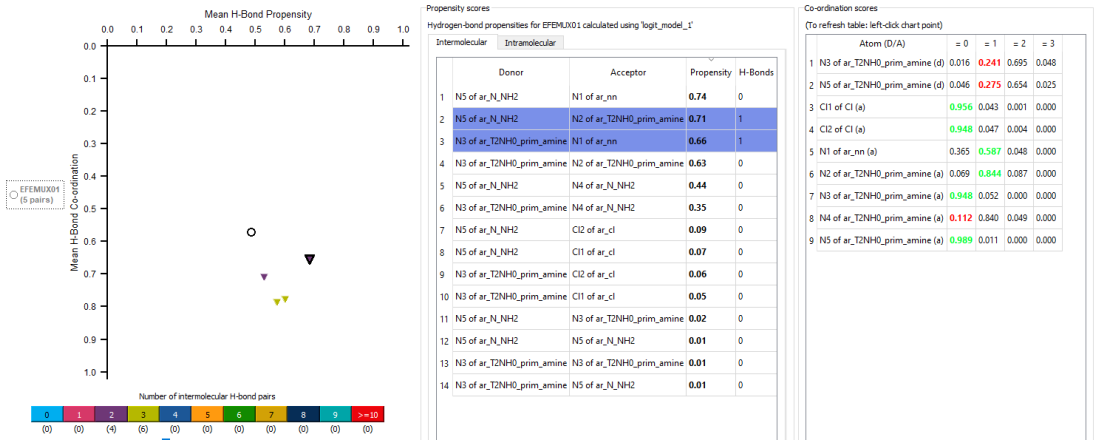


Figure 14; Coordination vs propensity plot, coordination scores and propensity scores as calculated in the HBP model for EFEMUX01 lamotrigine. Highlighted are the coordination scores and propensity scores for the network which produces the highest average coordination score (yellow triangle). mages generated using Mercury 2023.2.0 (Build 382240), CSD v5.41, using the Hydrogen Bond Propensities feature located within the "Polymorph Assessment" section of the "CSD-Materials" tab.

The propensity score for a given donor/acceptor pair represents the statistical likelihood of a hydrogen-bond forming between the two. A high propensity score suggests that a favourable interaction forms between the donor/acceptor pair which is likely a stabilising interaction. Low propensity scores indicate hydrogen-bond donor/acceptor pairings which are less common, suggesting a less-favourable interaction and lower stabilising effect on the crystal structure. Higher average propensity scores suggest a network of strong hydrogen-bond pairs which are most likely to form.

The average propensity score for EFEMUX01 is 0.573, suggesting that, on average, the hydrogen-bond pairings within the structure are favourable. However, as with the coordination scores, the HBP model predicts that there are several potential hydrogen-bonding networks which can be formed by lamotrigine which utilise more favourable hydrogen-bond donor/acceptor pairings on average. The HBP model predicts that a network utilising two intermolecular hydrogen-bonds, between N5 of ar_N_NH2 and N2 of ar_T2NH0_prim_amine and N3 of ar_T2NH0_prim_amine and N1 of ar_nn would be possible, and would result in a more stabilising network and more satisfied hydrogen-bonding partners overall.



	Donor	Acceptor	Propensity	H-Bonds
1	N5 of ar_N_NH2	N1 of ar_nn	**0.74**	0
2	N5 of ar_N_NH2	N2 of ar_T2NH0_prim_amine	**0.71**	1
3	N3 of ar_T2NH0_prim_amine	N1 of ar_nn	**0.66**	1
4	N3 of ar_T2NH0_prim_amine	N2 of ar_T2NH0_prim_amine	**0.63**	0
5	N5 of ar_N_NH2	N4 of ar_N_NH2	**0.44**	0
6	N3 of ar_T2NH0_prim_amine	N4 of ar_N_NH2	**0.35**	0
7	N5 of ar_N_NH2	Cl2 of ar_cl	**0.09**	0
8	N5 of ar_N_NH2	Cl1 of ar_cl	**0.07**	0
9	N3 of ar_T2NH0_prim_amine	Cl2 of ar_cl	**0.06**	0
10	N3 of ar_T2NH0_prim_amine	Cl1 of ar_cl	**0.05**	0
11	N5 of ar_N_NH2	N3 of ar_T2NH0_prim_amine	**0.02**	0
12	N5 of ar_N_NH2	N5 of ar_N_NH2	**0.01**	0
13	N3 of ar_T2NH0_prim_amine	N3 of ar_T2NH0_prim_amine	**0.01**	0
14	N3 of ar_T2NH0_prim_amine	N5 of ar_N_NH2	**0.01**	0

	Atom (D/A)	≥ 0	≥ 1	≥ 2	≥ 3
1	N3 of ar_T2NH0_prim_amine (d)	0.016	0.241	0.695	0.048
2	N5 of ar_T2NH0_prim_amine (d)	0.046	0.275	0.654	0.025
3	Cl1 of Cl (a)	0.956	0.043	0.001	0.000
4	Cl2 of Cl (a)	0.948	0.047	0.004	0.000
5	N1 of ar_nn (a)	0.365	0.587	0.048	0.000
6	N2 of ar_T2NH0_prim_amine (a)	0.069	0.844	0.087	0.000
7	N3 of ar_T2NH0_prim_amine (a)	0.948	0.052	0.000	0.000
8	N4 of ar_T2NH0_prim_amine (a)	0.112	0.840	0.049	0.000
9	N5 of ar_T2NH0_prim_amine (a)	0.989	0.011	0.000	0.000

Figure 15; Coordination vs propensity plot, coordination scores and propensity scores as calculated in the HBP model for EFEMUX01 lamotrigine. Highlighted are the coordination scores and propensity scores for the network which produces the highest average propensity score (purple triangle). mages generated using Mercury 2023.2.0 (Build 382240), CSD v5.41, using the Hydrogen Bond Propensities feature located within the "Polymorph Assessment" section of the "CSD-Materials" tab.

In summary, HBP analysis of EFEMUX01 lamotrigine suggests that there are multiple alternate hydrogen-bonding networks its molecules could adopt. The hydrogen-bonding network in EFEMUX01 utilises five intermolecular hydrogen-bonds, with several bifurcated donor and acceptor atoms. This results in a weaker overall network, forming several unusual or unfavourable hydrogen-bonds. Subjecting EFEMUX01 to a polymorph screen would be likely to yield alternative polymorphs, as alternate forms could form more favourable and stronger hydrogen-bonding networks using fewer hydrogen-bonding partners, as predicted by the HBP.

1.2.2 Conclusions

SFI has been used to evaluate the intramolecular, intermolecular and supramolecular environment of lamotrigine. The informatics-based approach provided rapid evaluation of the different length scales from within the lamotrigine crystal, and help to provide rationale for the observed lack of polymorphism. Furthermore, the advancements made to the tools within the SFI workflow since the publication of the 2011 study have been highlighted, namely FIMs and HBP analysis.

A further test of the power of SFI would be the investigation of a polymorphic compound, in order to evaluate the ability of the SFI workflow to rationalise observed differences in polymorphic stability through analysis of the intramolecular, intermolecular and supramolecular environments observed within the crystals of each polymorph. This analysis would demonstrate the ability of the SFI workflow to understand and make predictions regarding polymorphic stability,

and hence could make SFI a rapid and computationally inexpensive addition to polymorph screening practises.

1.3 Polymorphophores

It has been long accepted within the pharmaceutical field that some structural motifs are privileged with respect to their biological activity. Such compounds are referred to as *pharmacophores,* the knowledge of which can be extremely helpful in speeding up the development of new active pharmaceutical models, particularly when the 3D structure of the binding site is unknown [53]. There exists a similar concept in the field of asymmetric transition metal catalysis, where certain ligands are preferred due to their compatibility with many different metal centres and conditions, whilst providing a consistent level of stereocontrol [54].These classes highlight that certain structural motifs can be privileged with respect to certain chemical effects, and these motifs can be powerful tools in their respective settings. Researchers have theorised there may be structural motifs which may possess similar properties with respect to a compound displaying polymorphism. For instance, there are multiple examples of pharmaceuticals which contain sulfonamides displaying polymorphism[55,56], though there are examples to the contrary[57,58].

When researchers were looking to investigate polymorphically privileged structures attention was naturally drawn to structures with a high number of experimentally accessible polymorphic forms. Scaffolds such as 5 methyl-2-[(2-nitro phenyl)amino]-3-thiophenecarbonitrile (ROY), carbamazepine, and fenamic acid derivatives are all compounds with a large number of experimentally accessible polymorphic forms. Chemical structures and number of characterised polymorphic forms for ROY, carbamazepine, flufenamic acid and tolfenamic acid can be found in Figure 16.

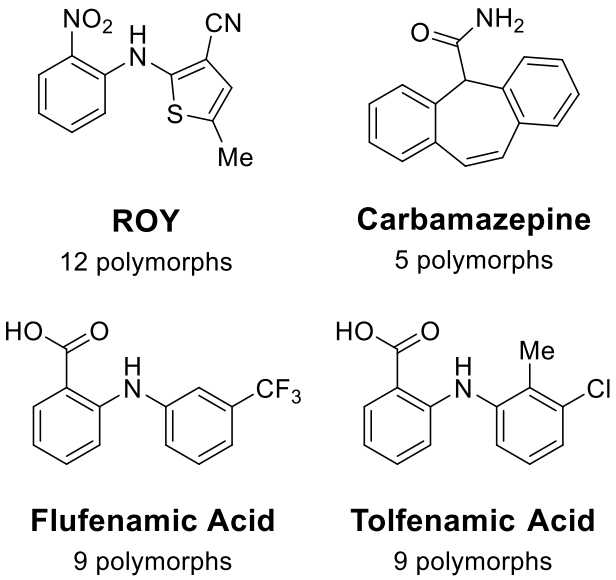


Figure 16; chemical structures and number of structurally characterised polymorphic forms for ROY, carbamazepine, flufenamic acid and tolfenamic acid.

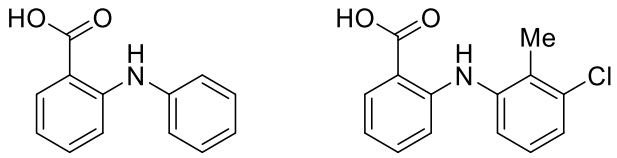
Moreover, further investigation into some of these motifs demonstrated that polymorphism is also observed in their structural derivatives[59–65]. This led to the coining of the term *polymorphophore* by Matzger in 2008, described as "a structural element that, when incorporated into a molecule, favours the formation of polymorphic forms". Since the coining of the term, a number of studies have been conducted into various potential polymorphophores, in order to investigate the structure-polymorphism relationship of the polymorphophore and its derivatives.

1.3.1 Fenamates

Fenamates are a form of N-phenyl-substituted anthranilic acid derivatives commonly used as non-steroid anti-inflammatory drugs (NSAID) in pain relief[66]. The fenamate structure was one of the original group of polymorphophores proposed by Matzger, and today exists as one of the most thoroughly investigated polymorphophores both computationally and experimentally.

1.3.1.1 Uzoh Study, 2012

The first investigation into the fenamate polymorphophore was performed by Uzoh et al, in 2012[67]. This study involved a computational investigation into two fenamate derivatives with differing polymorphic profiles: fenamic acid (**FA**), which is known to display one polymorph, and tolfenamic acid (**TA**), which is known to display 8 polymorphs. Structural diagrams and number of polymorphs for **FA** and **TA** can be found in Figure 17.



FA
Fenamic Acid
1 polymorph

TA
Tolfenamic Acid
9 polymorphs

Figure 17; Chemical structures and number of experimentally observed forms/polymorphs for fenamic acid (FA) and tolfenamic acid (TA)

The study by Uzoh et al. produced a crystal energy landscape for both **FA** and **TA** which was used to explore the similarities and differences in packing motifs observed between the lowest energy forms of both compounds. These crystal energy landscapes revealed a significant energy gap between the lowest-energy structures of **FA**, whilst **TA** displays a cluster of structures of similar energy around the energy minimum.

Analysis of the predicted structures located around the energy minimum for **FA** and **TA** reveals $R_2^2(8)$ dimer is present in all low energy structures of **FA** and **TA**. The differences observed between low energy structures of **FA** and **TA** are observed in the packing of the phenyl rings which protrude from either side of the roughly planar acid dimer building block. The torsional flexibility of the fenamate structure allows for close packing of the aromatic rings in **FA**. However, the bulky substituents present in **TA** prevent the efficient packing observed for the planar motif of **FA**. This ultimately results in a number of different packing arrangements for **TA** displaying similar energy, and hence resulting in the higher number of experimentally observed polymorphs of **TA**. Lattice energy versus packing coefficient graphs for the crystal energy landscapes calculated for **FA** and **TA** can be found in Figure 18 and Figure 19, respectively.

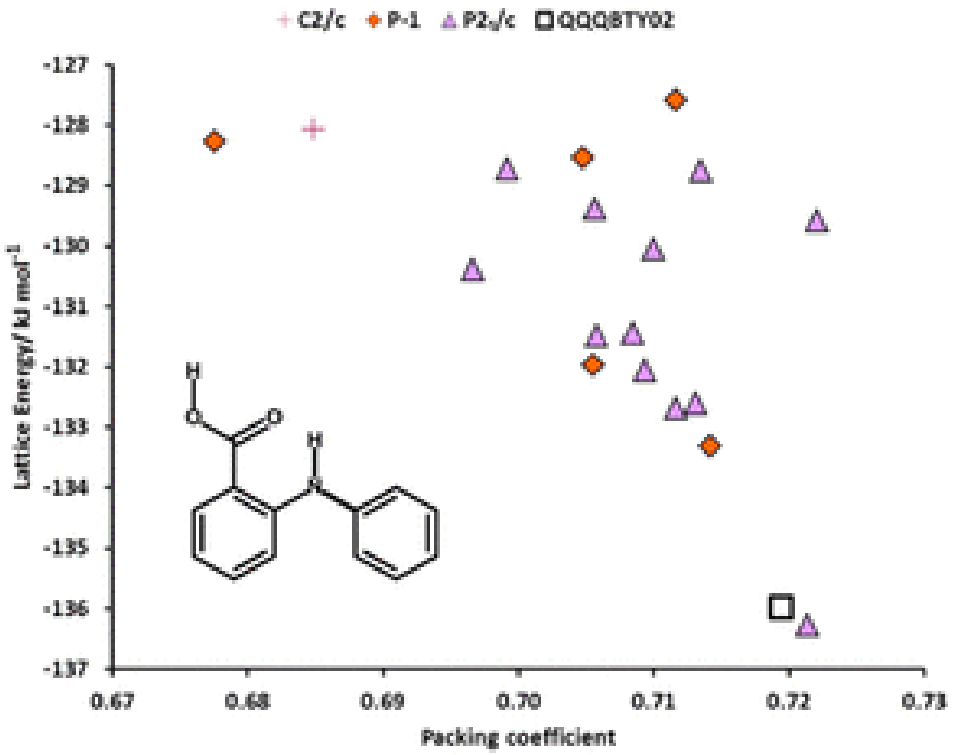


Figure 18; lattice energy vs packing coefficient for structures generated as Crystal energy landscape of fenamic acid (FA), experimentally observed structure is highlighted by a white box. Calculated structures with space groups C2/c, P-1 and P21/c are highlighted by pink crosses, orange diamonds and purple triangles, respectively.

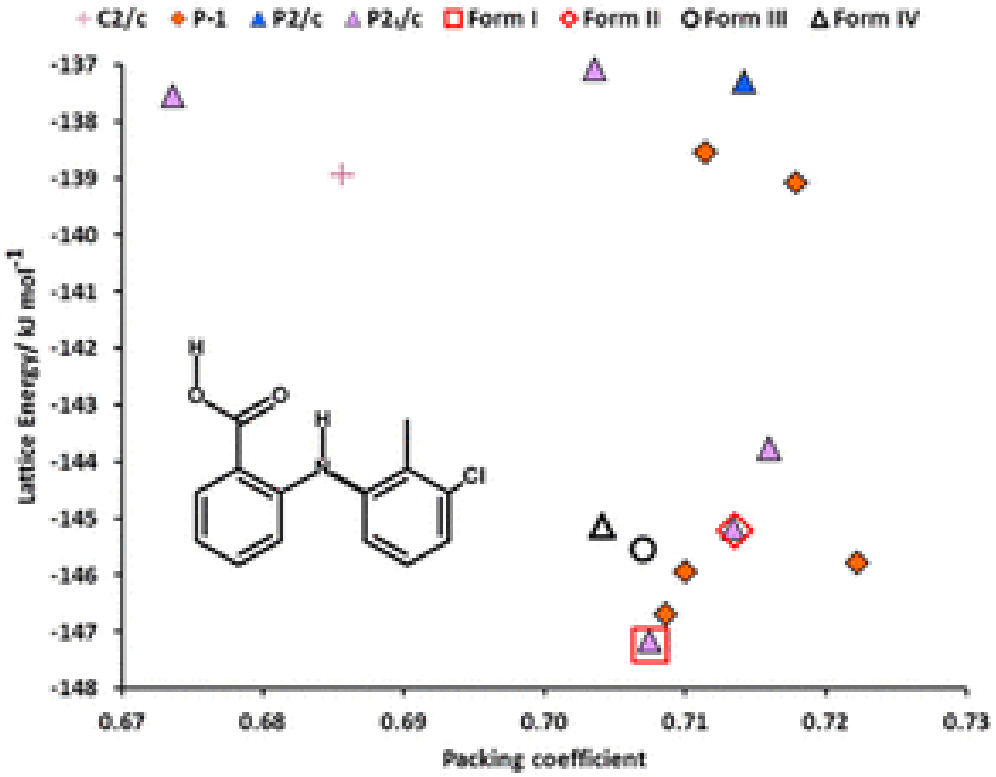


Figure 19; lattice energy vs packing coefficient for structures generated as Crystal energy landscape of tolfenamic acid (TA), experimentally observed Forms I – IV are highlighted by red box, red diamond, white circle and white triangle, respectively. Calculated structures with space groups C2/c, P-1, P2/c and P21/c are highlighted by pink cross, orange diamonds, blue triangles and purple triangles, respectively.

The study from Uzoh et al. revealed that it was the substituents of the fenamate structure which were the origin of its polymorphism, specifically the 2 and 3 positions of the noncarboxylated aromatic ring.

1.3.1.2 Lopez-Meijas Study, 2015

A 2015 study from Lopez-Meijas et al. looked to expand on these findings by subjecting a set of six close structural analogues of **TA** to a small-scale polymorph screen in order to investigate their SPR. These six structural analogues consisted of two commercially available fenamic acids, mefenamic acid (**MA**) and fenamic acid (**FA**), along with four structural analogues synthesised for this study, labelled LM1 – LM4. Structures for the six analogues screened as part of the 2015 study can be found in Figure 20.

The systematic approach to the selection of derivatives allowed for a thorough investigation into the 2 and 3 positions of the noncarboxylated ring, along with probing the importance of halogen interactions and steric bulk. The small scale PIHn screen revealed all forms except for **FA** to be polymorphic* (or at least to display multiple distinct forms), further demonstrating the importance of

substitution at the 2/3 positions of the noncarboxylated ring to the polymorphism of fenamate structures.

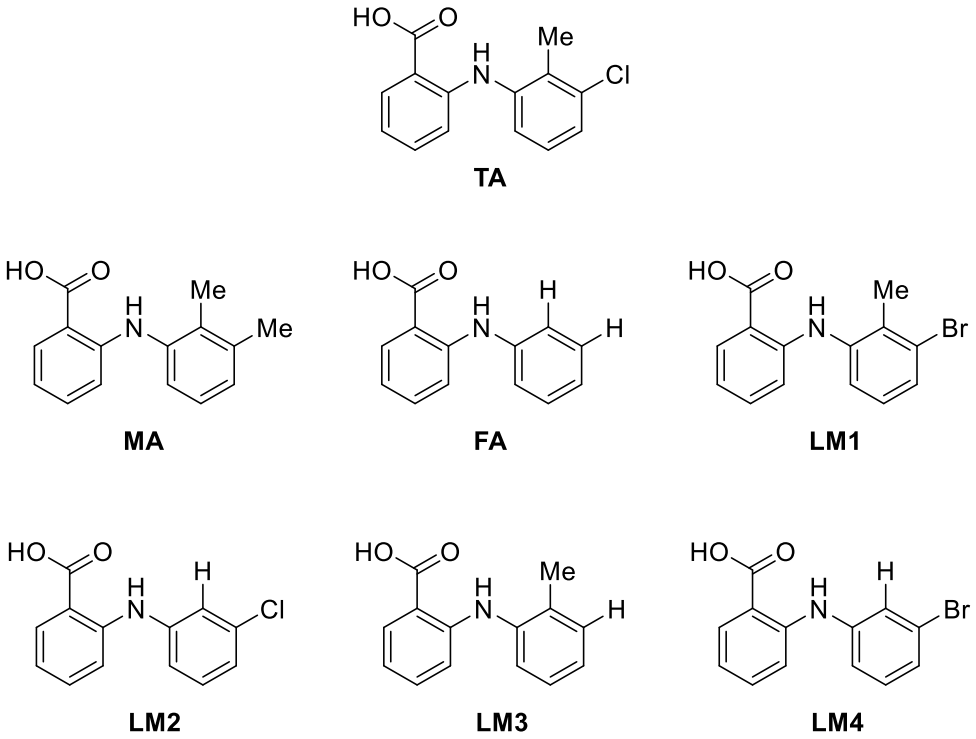


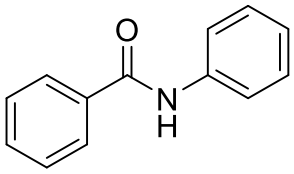
Figure 20; Chemical structure for tolfenamic acid (TA). Chemical structures of close structural derivatives mefenamic acid (MA), fenamic acid (FA) and LM1 – LM4, which were subject to a polymorph screen by Lopez-Meijas et al. in order to investigate the SPR of TA.

Computational analyses were focused on ascertaining isostructurality of the different polymorphs of the structural derivatives of **TA**, and ultimately revealed the importance of halogen – halogen interactions in the polymorphs of **TA** and its derivatives, along with the importance of the methyl group in the accessibility of the packing modes observed experimentally across forms of **TA**.

The Lopez-Meijas study into **TA** and its derivatives serves as the most thorough investigation into the SPR of the fenamate polymorphophore to date. The study built upon existing studies into the fenamate polymorphophore which identified the importance of substitution at the 2 and 3 positions of the noncarboxylated ring, and provided insight into the intermolecular factors which play a role in the polymorphism of fenamic acid derivatives, through investigation of derivatives which maintain the R22(8) carboxylic acid dimer observed across different polymorphic fenamic acid derivatives. Further studies could look to probe the effect of larger structural changes, positions 3 – 5 of noncarboxylated ring, or try to probe the dimer itself through the modification of H-bond donor/acceptor groups.

1.3.2 N-phenylbenzamides

Another polymorphophore which has been investigated experimentally is based around the N-phenylbenzamide (NPBA) scaffold. Chemical structures for the NPBA polymorphophore can be found in Figure 21. NPBA was identified by Matzger amongst the first proposed polymorphophore structures, and has since been subject to a number of experimental and computational investigations into its SPR.



NPBA
N-phenylbenzamide

Figure 21; chemical structure for N-phenylbenzamide (NPBA), structure labelled by Matzger in 2008 as a potential polymorphophore.

1.3.2.1 Zipp Study, 2013

The first study into the SPR of the NPBA polymorphophore was conducted in 2013 by Zipp et al., and focused on the synthesis, screening and computational investigation of four structural derivatives of the NPBA scaffold. These four structures were synthesised as part of a study into a biomimetic vitamin B12 model, resulting in NPBA derivatives with significant structural differences to the patent structure. Structures for the five NPBA derivatives investigated as part of the Zipp study can be found in Figure 22.

The polymorph screen performed during this study revealed that all five NPBA derivatives had at least two accessible polymorphic forms, an observation which could be seen as evidence for NPBA being a polymorphophore.

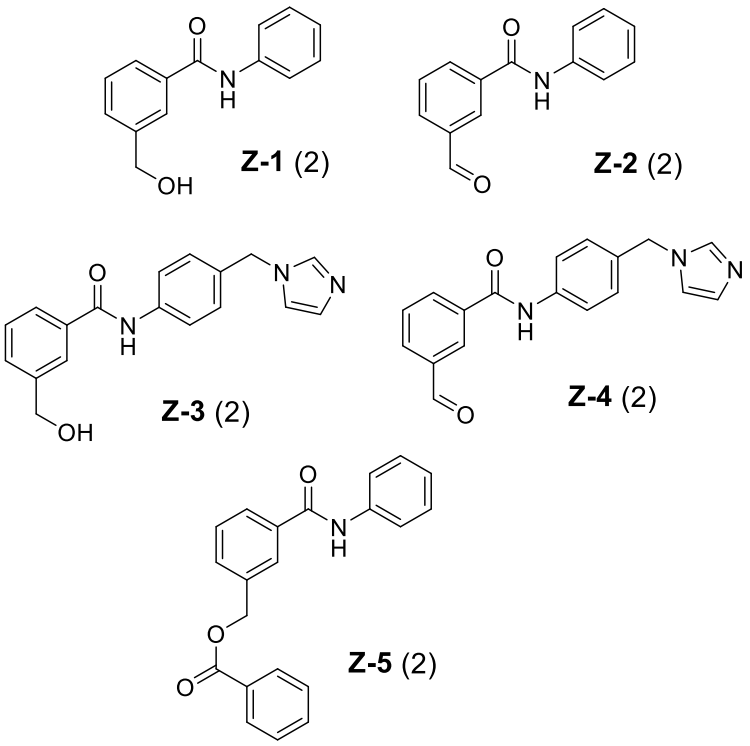


Figure 22; chemical structures of the structural derivatives of the NPBA polymorphophore synthesised as part of the 2013 investigation of the SPR of the NPBA scaffold. The number in backets highlights the number of polymorphic forms observed by Zipp et al.

The computational investigation into the SPR of the NPBA scaffold was performed via the use of DFT to calculate relative molecule – molecule interaction energies for each derivative, whilst dominant lattice interactions were calculated using PIXEL. Computational analysis found that a number of different OHO, OHN and NHO H-bonds were present as the dominant bonding motifs present within the crystal lattices of the polymorphic forms of Z-1 - Z-5. This could be due to the relatively large structural changes introduce to the NPBA scaffold in derivatives Z-1 – Z-5, introducing significant changes to the potential interactions within the lattice and making rationalisation of the impact of those changes on the polymorphic landscape difficult.

1.3.2.2 Bhandary Study, 2017

The 2017 study from Bhandary et al. was designed with the previous polymorphophore studies from Lopes-Meijas and Matzger et al. in mind, with smaller functional changed introduced to the NPBA scaffold in order to more closely map the changes in crystal properties. This study was focused around the polymorph screening and computational investigation of 30, a monomorphic NPBA derivative in which C-H the 3-position of the nitrogen-bound benzene ring is replaced with a C-C=C bond, and ten structural derivatives of 30 featuring modifications to the carbon-bound benzene ring. Chemical structures for 30 and

the ten other NPBA-derivatives investigated by Bhandary et al. can be found in Figure 23.

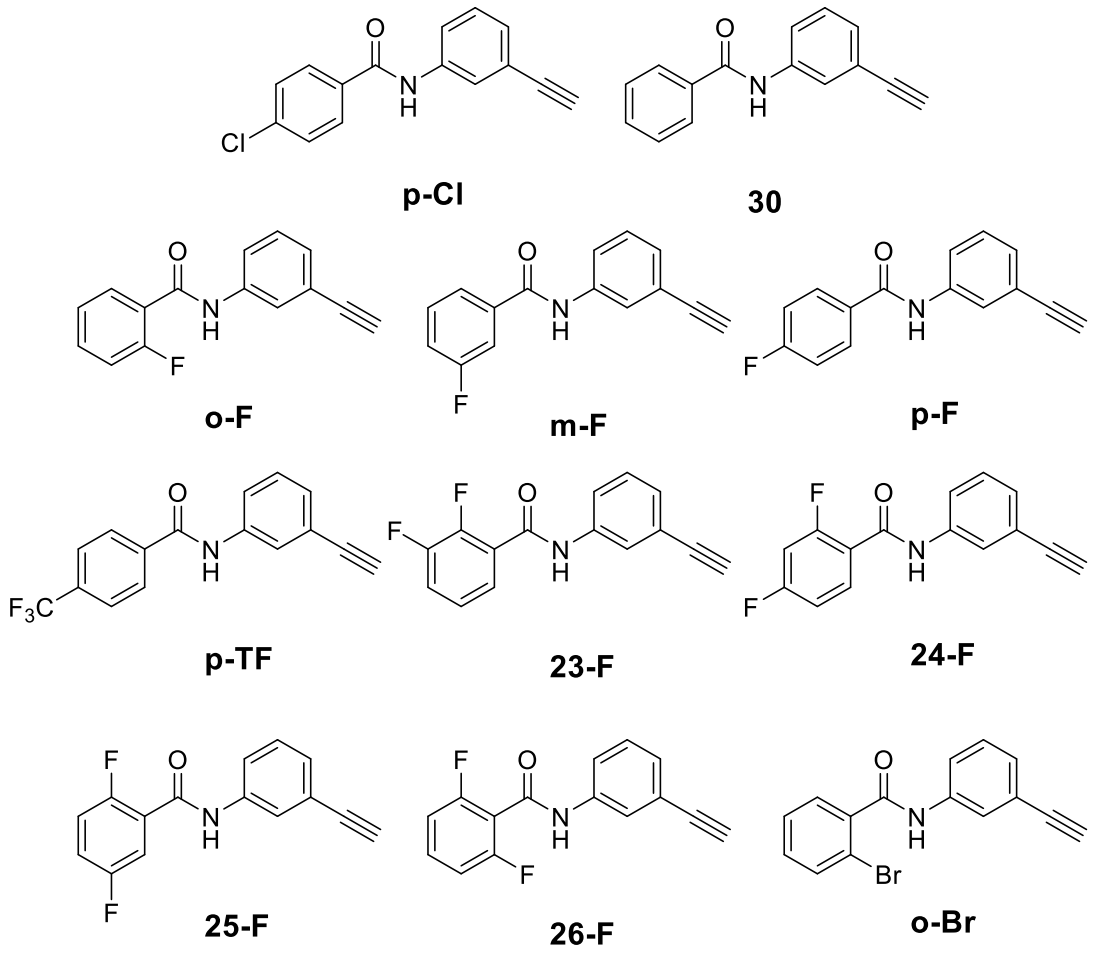


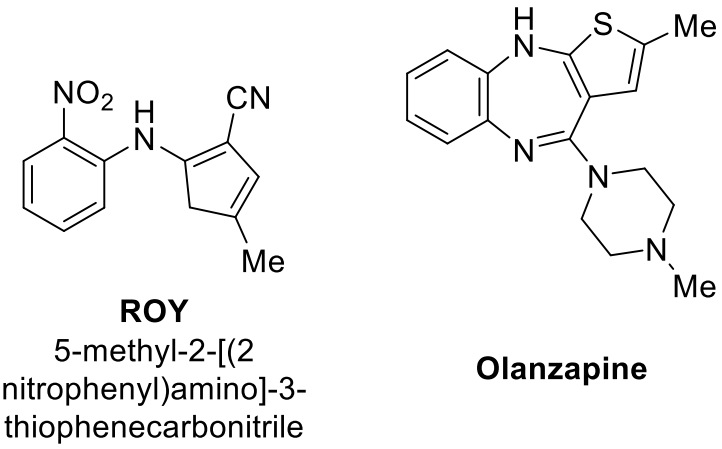
Figure 23; the library of NPBA derivatives which were screened by Bhandary et al. in their investigation into the SPR of the NPBA polymorphophore.

The investigation found that all derivatives except for 30 were found to display polymorphism. Computational analysis consisted of investigation of molecular conformations, synthon preferences and the generation of an interaction topology map via the use of DFT calculations. These analyses highlighted the importance of the N-H – O=C H-bonds to the stability of crystals containing the NPBA motif, however, it is secondary interactions involving substituents play a role in the accessibility of polymorphic forms, similar to the findings from the Lopes-Meijas study.

1.3.3 ROY

5-methyl-2-[(2-nitrophenyl)amino]-3-thiophene carbonitrile, otherwise known as ROY, due to its discovery in the form of red, orange and yellow crystals, was originally synthesised as a precursor to the antipsychotic medication,

olanzapine[68]. Chemical structures for ROY and olanzapine can be found in Figure 24.



ROY
5-methyl-2-[(2 nitrophenyl)amino]-3-thiophenecarbonitrile

Olanzapine

Figure 24; chemical structures for ROY and olanzapine

The unique colour profile of ROY and its polymorphs, alongside the ease with which multiple polymorphic forms were discovered, made ROY an intriguing structure for the study of polymorphism as a phenomenon.

1.3.3.1 Polymorphism of ROY

In 2001, Yu et al. sought to isolate and characterise the known polymorphs of ROY using crystallographic and thermal methods[69]. This led to the characterisation of the first six polymorphic forms of ROY, these being: orange needle (**ON**), orange prism (**O**), red prism (**R**), yellow needle (**YN**), yellow prism (**Y**) and orange-red plates (**ORP**). Images of the six initially discovered polymorphs of ROY can be found in Figure 25.

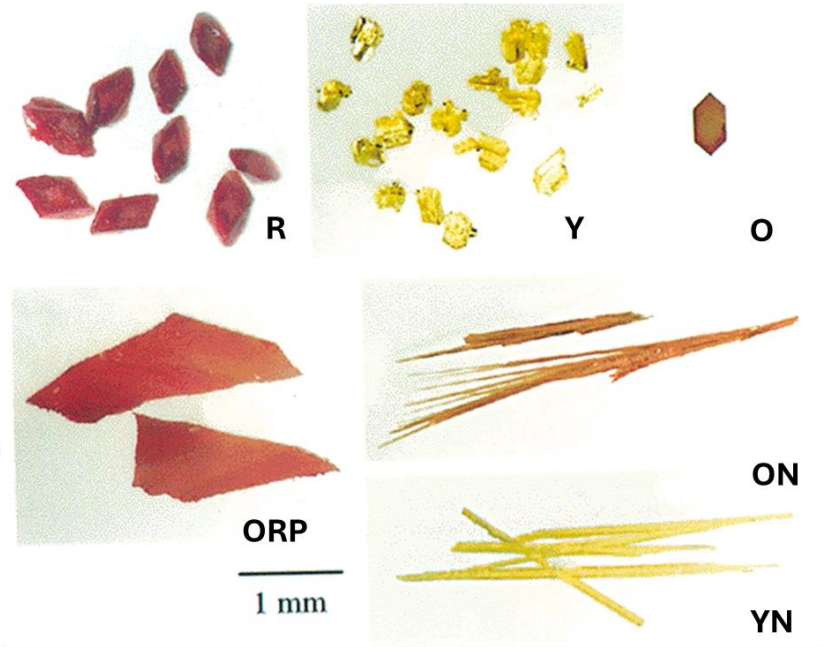


Figure 25; images of crystals of the first six ROY polymorphs of ROY discovered: red prism (R), yellow prism (Y), orange prism (O), orange-red plate (ORP), orange needle (ON), and yellow needle (YN). Reprinted (adapted) with permission from [69]. Copyright 2001 American Chemical Society.

This study also provided the first investigation into the polymorphic stability of the ROY polymorphs, utilising Differential Scanning Calorimetry (DSC) Yu et al. were able to ascertain melting points for **R** (M.P. = 106.2 °C), **ON** (M.P. = 114.8 °C), O (M.P. = 112.7 °C) and **Y** (M.P. = 109.8 °C). Thermal data for **ORP** and **YN** weren't obtained due to thermal instability. Slurrying experiments revealed **Y** to be the most stable form between 20 – 60 °C, but **ON** and **O** forms are both more stable than **Y** at elevated temperatures, displaying an enantiotropic relationship. Polymorph **R** was found to be less stable than **Y** at temperatures above RT. Melt grown crystals of **YN** were found to convert to **Y**, **ON** or **O** without discrimination at RT, in hours or days. **ORP** was found to undergo slow conversion to **Y** at RT. The variance in polymorphic stability, along with the enantiotropic relationship between **Y** and **ON**/**O**, is seemingly connected to the different conformations adopted by ROY molecules within the different polymorphs, leading to variation in structure entropy along with intermolecular environment.

These initial studies led to increased interest in ROY and its clear connection to the phenomenon of polymorphism, and an additional six unique polymorphic forms have been identified and fully characterised since then. The first two additional polymorphs were originally reported in 2005 by Chen et al.[70], when they observed a yellow ROY forms displaying a unique Raman spectrum to **YN** and **Y** during a cross-nucleation experiment[71,72]. This led to the discovery of the

Y04 form of ROY, though its metastability meant full characterisation of **Y04** would remain elusive until 2020.

The **Y04** Form of ROY was first observed from melt crystallisation of ROY at 130 °C encapsulated between two cover glasses. Upon cooling to 22 °C the distinct yellow Form of **Y04** was identified, but upon heating **Y04** rapidly converted to the **R** Form of ROY. Holding the temperature at 22 °C, however, led to the transformation of **Y04** to another, duller yellow form which also displayed a distinct Raman spectrum from **Y04**, **YN** and **Y**, dubbed the **YT04** Form of ROY. Images of the **R**, **Y04** and **YT04** Forms of ROY from the 2005 study can be found in Figure 26.

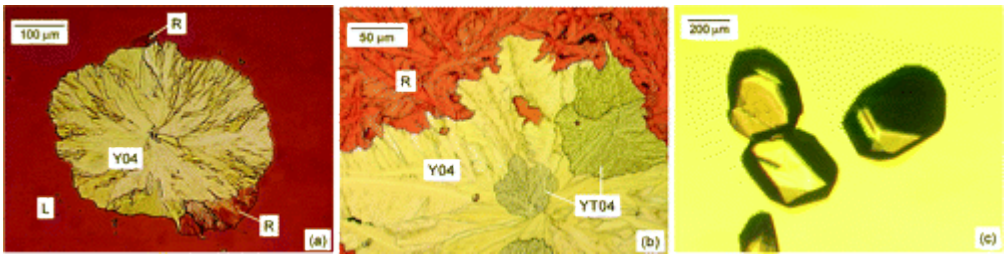


Figure 26; (Left) image of ROY Y04 converting to R. (Middle) a sample of ROY displaying R, Y04 and YT04 crystals. (Right) Isolated single crystals of YT04. Image taken from the 2005 study from Chen et al[70]. Reprinted (adapted) with permission from [70]. Copyright 2005 American Chemical Society."

The **YT04** Form of ROY was stable enough to collect both crystallographic and thermal data. These data revealed **YT04** to be less stable than **Y**, **O** and **ON** at high temperatures, but more stable than **R** and **YN**. However, ΔG versus temperature graph suggests that **YT04** would have an enantiotropic relationship with both **O** and **ON**, similar to the **Y** Form of ROY.

There was a relative lull in the discovery of new ROY polymorphs until 2018 when Tan and coworkers utilised a synergistic combination of CSP along with PXRD and Raman spectroscopy to identify the **R05** Form of ROY[73]. **RO5** was generated through the use of thin (< 5 um) films of ROY melt deposited on to cover glasses. Spontaneous crystallisation at room temperature yielded **R05** (along with several other ROY Forms). An image of **R05** from the 2018 study can be found in Figure 27.

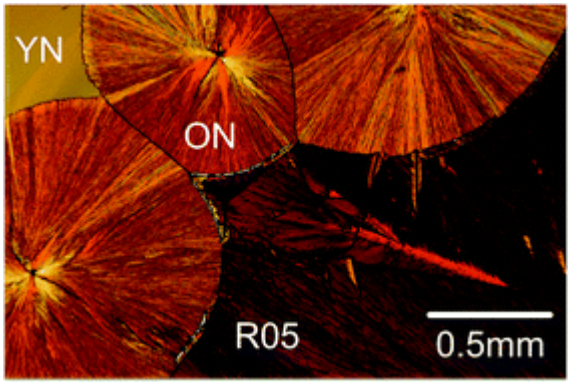


Figure 27; Image of crystals of the R05 Form of ROY alongside crystals of ON and YN. Image taken from the 2018 study from Tan et al[73]. Used with permission of the rights holder.

The **R05** form of ROY was the first ROY polymorph to be discovered for which Z' > 1. Upon standing **R05** would rapidly convert to **R** or **Y** if any of these forms were present, but, isolated **R05** could have a lifetime up to weeks.

In 2019 another ROY form was discovered, this time through the use of variable temperature PXRD. The **PO13** form of ROY was discovered by Gushurst et al. after a sample of the **YN** form of ROY was melted at 131 °C, before being cooled to 30 °C[74]. PXRD patterns associated with **PO13** were noted at 60 °C during the cooling step. An image of a mixed **PO13**/**Y** sample of ROY can be found in Figure 28.

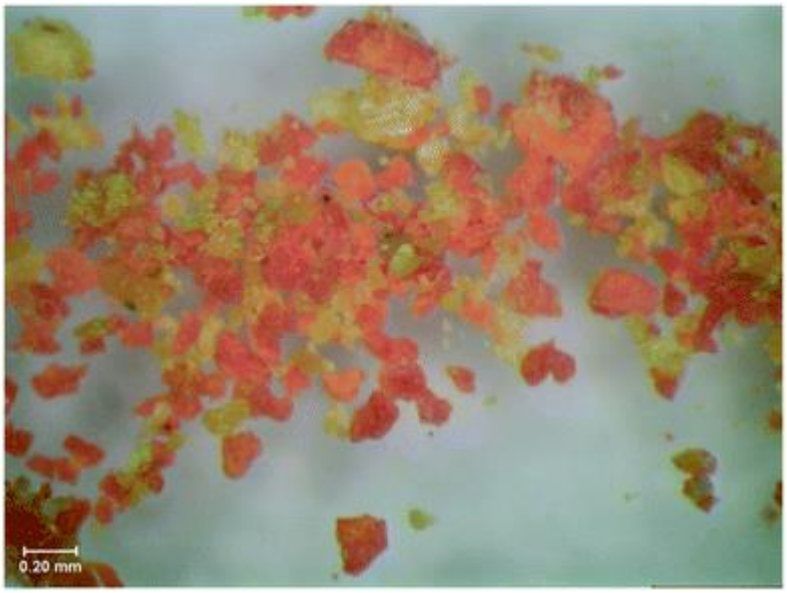


Figure 28; a mixed sample of Y and PO13 form of ROY taken after 3 months in storage. Image taken from the 2019 study from Gushurst et al [74]. Used with permission from the rights holder.

Structural determination for **PO13** was performed via comparison of experimental PXRD patterns to those obtained from CSP data. Thermal analysis was

performed on the **PO13** form of ROY, however thermal events for **PO13** were not observed on the thermogram.

Finally, the three most recent ROY polymorphs to structurally characterised were all reported in the summer of 2020. The first of which was the **Y04** form discussed above, originally reported by Chen et al., which was finally characterised 15 years after its initial discovery. Li et al.[25] utilised melt microdroplets of 200 – 400 um diameter in order to isolate single crystals of **Y04** for structural determination. An image of a crystal of the **Y04** form of ROY can be found in Figure 29.

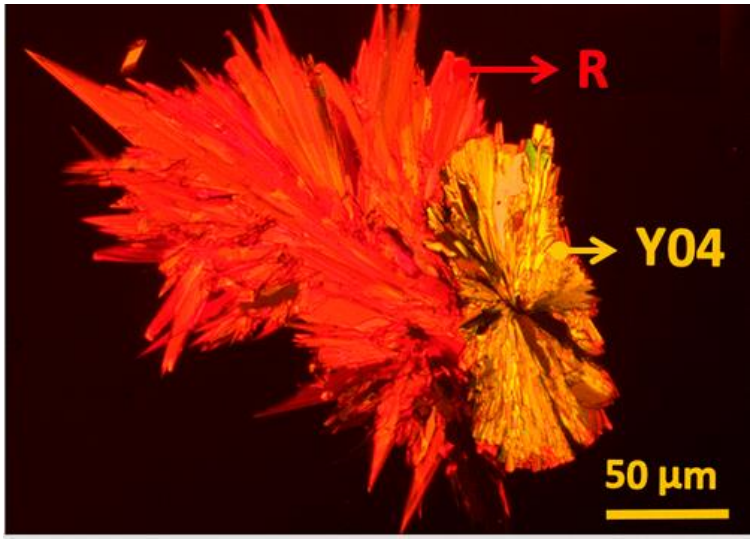


Figure 29; Image of the crystals of the R and Y04 Forms of ROY. Image taken from the 2020 study from Li et al. Reprinted (adapted) with permission from [25] Copyright 2020 American Chemical Society.

The report from Li et al. contained no thermal data, but observations made in the 2005 study by Chen et al. suggest the **Y04** form to be metastable, due to its rapid conversion to **YT04** upon standing[70].

The **R18** and **Y19** forms of ROY are the most recent polymorphs of ROY to be discovered, and each required the use of a novel crystallisation method in order to access and characterise each form. The **R18** form was discovered by Tyler et al. who utilised Encapsulated Nanodroplet Crystallisation (ENaCt) to perform a polymorph screen for ROY, during which the new form was discovered[75]. An image of a crystal of the **R18** form of ROY can be found in Figure 30.

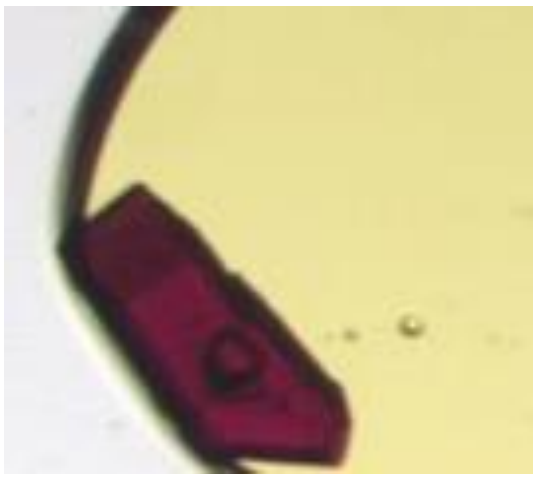
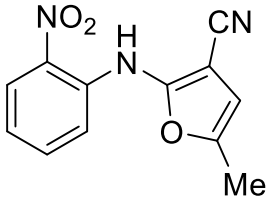


Figure 30; Single crystal of the R18 Form of ROY, generated and isolated using the ENaCt crystallisation method. Image taken from the 2020 study from Tyler et al[75]. Reprinted (adapted) with permission from [75]. Copyright 2019 American Chemical Society."

The **R18** polymorphs is only the second ROY form discovered to date for which Z' > 1 (along with the **R05** form). No thermal or polymorphic stability data were reported for **R18**.

he **Y19** From of ROY is, at the time of writing, the twelfth and latest ROY polymorph to be discovered[24]. Levesque et al. employed the use of mixed crystals, otherwise known as solid solutions[76], of ROY and a structurally related compound known as **FuROY** (a furan ROY derivative). The structure of **FuROY** can be found in Figure 31.



FuROY

Figure 31; chemical structure for FuROY, a furan derivative of the ROY scaffold used in the discovery of the Y19 form of ROY.

In order to isolate **Y19**, small amounts of 40:60 ROY:FuROY mixed crystals were added to a vigorously stirred suspension of supercooled molten ROY in water at 70 ^{o}C, yielding a polycrystalline solid, along with small amounts of **Y19**. The **Y19** form of ROY was found to convert to the **Y** form in hours or upon exposure to elevated temperatures or pressures. No images of **Y19** were collected as a part of this study.

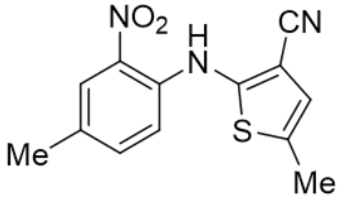
The vast number of experimentally accessible polymorphs of ROY, along with the varied and complex nature of the stabilities of the ROY polymorphs, makes ROY stand out as a scaffold of interest when it comes to investigation of polymorphism

Table 5; summarising crystallisation and stability data for the 12 ROY polymorphs

Form	Colour	Habit	Crystallisation Method	CSD Refcode	Stability
ON	Orange	Needle	Ambient crystallisation	QAXMEH	M.P = 114.8 °C
Y	Yellow	Prism	Ambient crystallisation	QAXMEH01	M.P = 109.8 °C
R	Red	Prism	Ambient crystallisation	QAXMEH02	M.P = 106.2 oC
O	Orange	Prism	Ambient crystallisation	QAXMEH03	M.P = 112.7 °C
YN	Yellow	Needle	Ambient crystallisation	QAXMEH04	M.P = 99.0 °C
ORP	Orange-red	Plate	Ambient crystallisation	QAXMEH05	M.P = 97.4 °C
YT04	Yellow	Prism	Encapsulated melt crystallisation	QAXMEH12	M.P = 106.9 °C
R05	Red	Prism	Thin film crystallisation	QAXMEH31	Polymorph conversion before melting
PO13	Pumpkin orange	Prism	Melting **YN**, cooling crystallisation	QAXMEH52	Polymorph conversion before melting
Y04	Yellow	Prism	Melt microdroplet crystallisation	QAXMEH53	Polymorph conversion before melting
R18	Red	Prism	ENaCt	QAXMEH57	Polymorph conversion before melting
Y19	Yellow	Prism	Mixed crystallisation	QAXMEH60	Polymorph conversion before melting

1.3.3.2 Polymorphism of ROY Derivatives

There have also been a number of studies focused on investigating the polymorphic landscapes of compounds structurally related to ROY. The first of which was performed by He et al in 2001, in which a ROY derivative was synthesised with the C-H bond at the 4'-position of the nitrobenzene ring was replaced by a C-CH_3 bond (herein referred to as **4'-methyl-ROY**), during an investigation into the conformational polymorphism of the ROY motif[63]. Four polymorphic forms of **4'methyl-ROY** were identified (red, dark red, light red and orange), but only three (red, dark red and light red) had their structures determined. Chemical structure of **4'-methyl-ROY** be found in Figure 32.

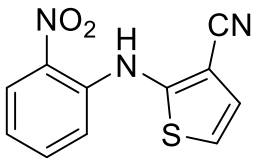


4-methyl-ROY

Four polymorphs

Figure 32; chemical structure for 4-methyl-ROY, along with the number of polymorphs observed in the 2001 study by He et al[63].

The next structural derivative of ROY to be investigated was produced by Li et al. in 2006. Similar to the 2001 study, Li et al. generated another ROY derivative in order to probe the conformational polymorphism of the ROY motif, this time generating a ROY derivative in which the C-CH_3 bond at the 5-poisition of the thiophene ring is replaced by a C-H bond (herein referred to as **demethyl-ROY**)[62]. Two polymorphs of **demethyl-ROY** were characterised, red and orange forms, whilst a third, yellow form was observed, but was identified as a disappearing polymorph. Chemical structure for **demethyl-ROY** can be found in Figure 33.



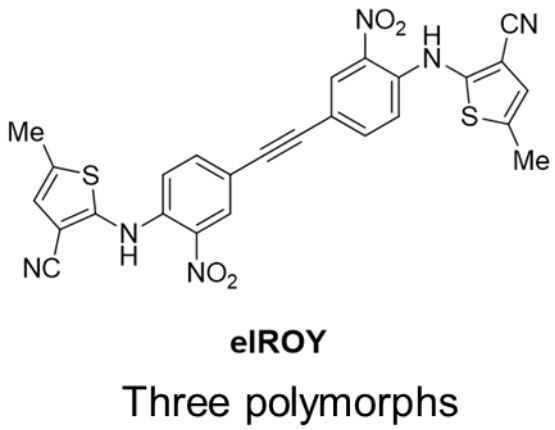
demethyl-ROY

Two polymorphs

Figure 33 Chemical structure for demethyl-ROY along with the number of polymorphs observed in the 2006 study by Li et al[62].

It was the polymorphism of these structural derivatives of ROY, alongside the unusually large number of polymorphic forms of ROY, which led Matzger to coin the term polymorphophore[77].

In a 2008 study, Matzger et al. sought to explore the concept of a polymorphophore using ROY as an example. In this study, a dimeric ROY derivative connected by an ethynyl bridge connected to the 4 position of the nitrobenzene ring of ROY, dubbed **elROY**, was designed, synthesised and subject to a polymorph screen. This polymorph screen identified three distinct forms of **elROY**, and seemed to confirm the ability of the ROY polymorphophore to increase the likelihood of polymorphism to occur. Chemical structure for **elROY** can be found in Figure 34.



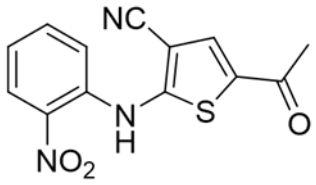
elROY

Three polymorphs

Figure 34; Chemical structure for el-ROY along with the number of polymorphs observed in the 2008 study by Lutker et al[77].

After this, there was a period of inactivity with respect to investigations into the SPR of the ROY polymorphophore, until the publication of a pair of studies from Nogueira et al. in 2022. These focus and construction of these studies was more in-line with studies of the fenamates and N-phenylbenzamides from the 2010's, than the previous studies of **4'-methyl-ROY** and **dimethyl-ROY**; signifying a shift in focus from the conformational polymorphism of ROY to an investigation of the ROY polymorphophore.

The first study was focused on a derivative of the ROY scaffold in which the methyl group at the 5 position of the thiophene ring of ROY was replaced by an acetyl group, producing the **5-acetyl-ROY** derivative[64]. Three polymorphic forms of **5-acetyl-ROY** were identified, burgundy plates (**5-acetyl-ROY 1**), orange needles (**5-acetyl-ROY 2**) and orange-yellow needles (**5-acetyl-ROY 3**), though full structural determination was only possible for **5-acetyl-ROY 1** and **5-acetyl-ROY 2**. Chemical structure for **5-acetyl-ROY** can be found in Figure 35.



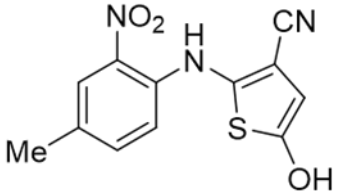
5-acetyl-ROY

Three polymorphs

Figure 35; Chemical structure for 5-acetyl-ROY, along with the number of polymorphs discovered in the 2022 study by Nogueira et al[64].

Computational analysis consisted of conformational analysis performed using DFT methods, and analysis of the dominant packing motifs observed within the crystal through the use of Hirschfeld surfaces, revealing that both 1 and 2 show different dominant packing motifs.

The most recent study into the polymorphism of ROY derivatives was a follow-on from the **5-acetyl-ROY** investigation, in which the same principles of synthesis, screening and computational analysis were employed to a structural derivative of **5-acetyl-ROY** in which the acetyl group attached to the thiophene ring is replaced by a primary alcohol, dubbed **ROY-ol**[65]. Four colour polymorphs of **ROY-ol** were observed by Nogueira et al., however only two were solved via single crystal XRD; **Polymorph 1** and **Polymorphs 2**. Chemical structure for **ROY-ol** can be found in Figure 36.



ROY-ol

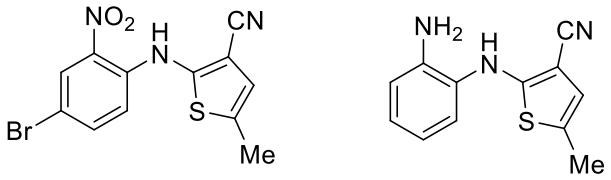
Two polymorphs

Figure 36; Chemical structure for ROY-ol, along with the number of polymorphs discovered in the 2022 study by Nogueira et al[65].

Hirschfeld surface analysis of Polymorph 1 and Polymorph 2 revealed that the same packing motif, an O-H -- N H-bond between the cyano nitrogen and alcohol proton of neighbouring **ROY-ol** molecules, is dominant in each form.

Along with the polymorphic ROY derivatives mentioned above, there are examples of ROY derivatives for which no polymorphic forms have yet been found. These include **BrROY,** a ROY derivative in which the C-H bond at the 4'-position of the nitrobenzene ring of ROY was replaced by a C-Br bond, which was

synthesised and screened as part of the 2008 study into **eI-ROY**. Another example would be a derivative reported by Leyva-Perez et al. in 2010. They reported an industrial route to the drug olanzapine which proceeds via ROY-like intermediate in which the nitro group is replaced with a primary amine, dubbed amino-ROY. Chemical structures for **BrROY** and **amino-ROY** can be found in Figure 37.



BrROY **amino-ROY**

Figure 37; chemical structures for the monomorphic ROY analogues BrROY (left) and amino-ROY (right).

Researchers found **BrROY** to display zero polymorphic forms despite extensive screening; whilst it in unclear whether **amino-ROY** was subject to any form of polymorph screen.

Chapter 2 Solid-Form Informatics Analysis of ROY Polymorphs from within the CSD

2.1 Aims and Objectives

The goal of this section was to explore the polymorphic landscape of the potential polymorphophore structure ROY using the Solid Form Informatics (SFI) workflow. The SFI workflow consists of a group of informatics tools constructed for use with the Cambridge Structural Database (CSD), a publicly available structural database containing the crystallographic information for over 1.2 million crystal structures.

The tools of the SFI workflow are used to evaluate the intra-, inter- and supramolecular environments present within the crystal structures of the 12 ROY polymorphs found within the CSD, compared to structures containing similar functionality from within the CSD. Comparison of these structural features of the 12 ROY polymorphs to trends in these features observed across structures found within the CSD will be used to highlight features of the ROY scaffold which may influence the high number of polymorphic forms accessible to the ROY scaffold.

The structural features highlighted through the SFI analysis performed in Chapter 2 will then be used to design a set of close structural analogues of ROY. These structural analogues will then be crystallised and characterised using single crystal X-ray diffraction (SC-XRD), and analysed using SFI to explore the influence of the structural modifications on the intra-, inter- and supramolecular environments of the ROY analogues. Finally, these ROY analogues will be subject to a low-/medium-throughput polymorph screen in order to ascertain the polymorphism of the ROY analogues, and hence evaluate the classification of ROY as a polymorphophore.

2.2 Introduction

In this section, SFI analysis of the ROY polymorphs found within the CSD is performed with the goal of establishing a Structure-Polymorphism Relationship (SPR) for the ROY polymorphophore. The informatics tools present within Mercury, along with the > 1.2 million structures of the Cambridge Structural Database (CSD), are utilised to explore the intramolecular, intermolecular and supramolecular environments of the 12 ROY polymorphs in an attempt to understand the source of the extreme polymorphism of ROY, along with rationalising their relative polymorphic stabilities.

Another key goal of this section is to provide critical analysis of the SFI workflow and its capacity to produce useful knowledge of the polymorphic landscape of the

ROY scaffold *in-silico*. There are many forms of ROY which have been characterised since its discovery at Eli Lily Laboratories in 1995[78], with a wide range of crystallisation techniques used to access these different forms[24,25,70,74]. During a typical polymorph screen, the goal is to find the optimal form to take forward further in production. For pharmaceutical products, this usually means finding the most thermodynamically stable form, such that this form can withstand the various conditions present during production, transportation and storage in a form which can still provide suitable therapeutic performance. If SFI is to be a useful tool for the design and implementation of polymorph screens, it would be useful for the SFI workflow to be able to discern between thermodynamically stable and less stable polymorphs, and hence be able to inform researchers when it is likely that more stable polymorphs may exist.

ROY was selected as the subject of this investigation for a number of reasons (notoriety, designation as a polymorphophore, druglike character etc.), but the primary reason was its extreme number of accessible polymorphic forms, 12, with ROY being the compound with the most solved single-component crystal structures within the CSD to date[79]. This makes ROY a rich source of diverse crystallographic data and also a suitable example to test the predictive powers of the SFI workflow. Moreover, as the compounds developed for the latter stages of this work are MMPs of ROY, the SFI analyses of these compounds should only bear minor differences to the SFI performed for the ROY polymorphs. This allows us to map the observed changes in the output of the SFI analysis to their respective changes to the chemical structure of the ROY scaffold. Coupling this with the results of the polymorph screen, this analysis could provide useful insight into the changes in the polymorphic landscape associated with the change in chemical structure, and hence a more thorough evaluation of the polymorphic landscape of a given derivative *in-silico*.

The available thermal data for the ROY polymorphs shows that the **ON** form is most stable at high temperatures, showing no conversions upon heating in a slurry at 90 – 100 ^{o}C. The **Y** form is the most stable at temperatures up to RT, displaying an enantiotropic relationship with **ON**. The **O** form is similar to **ON**, displaying an enantiotropic nature with both forms. The **YN** form was found to convert to **Y**, **ON** or **R** over hours to days upon standing, whilst heating slurries of **R** at 70 – 100 ^{o}C for hours to days resulted in conversion to **Y**, **O** or **ON**[69]. Thermal and DFT data acquired by Chen et al. in 2005 suggest **YT04** is the second most stable form at 0 K, behind **Y**, and the most dense polymorph to date[70]. Upon standing, the **R05** form converts to **Y** or **R** over a few hours[73].

The remaining polymorphic forms are significantly more metastable, rapidly converting to more stable forms and requiring unique crystallisation methods in

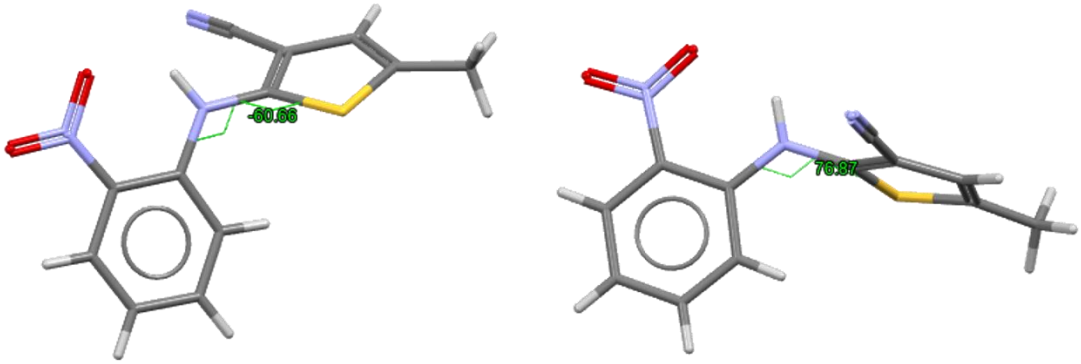
order to access a form suitable for characterisation. Gurhurst et al. in 2019 found the **PO13** form to rapidly convert to **Y** upon standing[74]. The **Y04** form was first discovered by Chen et al. in 2005, but its rapid conversion to **YT04** made characterisation elusive, until Li et al. were able to generate crystals of sufficient quality via melt microdroplet crystallisation[25]. Polymorphic conversion data for **R18** are not reported, but due to the encapsulated nanocrystallisation method used to access this form[75] it is thought to display similar stability to **Y04**. The **Y19** form is the latest form to be discovered to date, being discovered by Levesque et al. via mixed crystal seeding, and is found to rapidly convert to **Y** upon standing or gentle heating[24]. Based on the available thermal data, the relative polymorphic stability ranking for the twelve ROY polymorphs within the CSD is as follows: **ON** > **O** > **Y** > **YT04** > **R** > **ORP** = **YN** > **PO13** = **Y04** = **R18** = **Y19**.

In order to critically assess the power of the SFI workflow to guide decisions made during a polymorph screen, its ability to discern between *stable* polymorphs (i.e. **Y**, **ON**, **R**, **O** and **YT04**) and *metastable* polymorphs (i.e. **PO13**, **Y04**, **R18** and **Y19**) will be evaluated. Moreover, the accessibility of each step of the workflow will be assessed by examining how much chemical knowledge, software familiarity and experience using data-driven models are required to extract useful knowledge at each stage.

2.3 Intramolecular Environment

The intramolecular environment of molecules within different polymorphs of ROY dictates the type and strength of interactions which may be formed within each crystal, and hence play a role in the overall stability of a given polymorph. Intramolecular factors such as bond lengths, valence angles, molecular torsions and ring geometries can all vary between polymorphs and can have significant effects on the intermolecular interactions accessible within each crystal, and the supramolecular assemblies they aggregate to form. Moreover, any irregularities found within the crystal structure can suggest issues during structural determination, and hence identify unreliable data.

Mogul Geometry Check[80] (MGC) was used to evaluate the intramolecular environments present within the 12 unique ROY polymorphs within the CSD[25]. As demonstrated in Chapter 1.3, MGC is able to compare intramolecular factors such as bong length, valence angle, torsion angle and ring geometry from a desired structure with examples containing similar functionality from within the CSD. For the 12 ROY polymorphs it was found that the only intramolecular factor to vary significantly was the C-N-C-S torsion angle observed between the two aromatic rings, labelled as θ_{thio}.



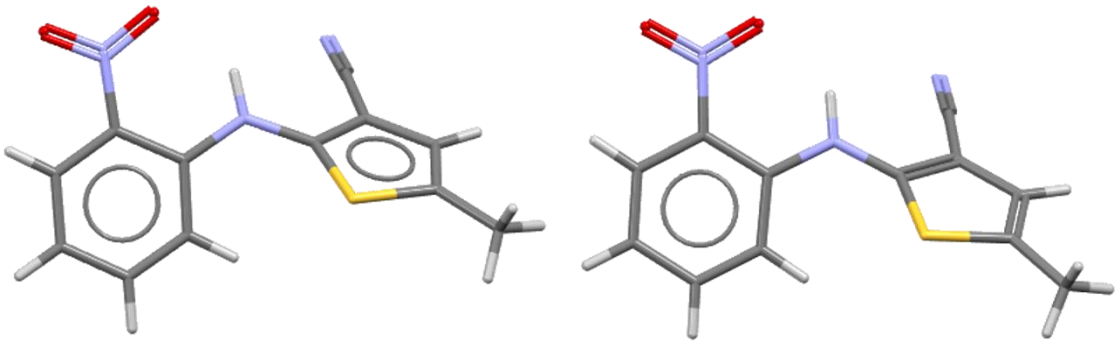
ON (QAXMEH)
θthio = 60.66°

Y04 (QAXMEH53)
θthio = 76.87°

Figure 38; (Left) θ_{thio} torsion for the ON (QAXMEH) form of ROY, and its corresponding value. (Right) θ_{thio} for the Y04 (QAXMEH53) form of ROY, and its corresponding value.

θ_{thio} could play an important role in the stabilisation of the ROY polymorphs due to the angle between the aromatic rings being related to the degree of overlap of the p-orbitals which make up the system, which has a stabilising effect on the structure as a whole. This, in turn, could play a role in determining the strength and type of intermolecular interactions formed by ROY in each of the different ROY forms. This makes θ_{thio} a useful metric to analyse when attempting to rationalise the differences in polymorphic stability observed for the ROY polymorphs.

Using MGC to evaluate θ_{thio} values for the 12 for polymorphs should allow for rapid comparison of this torsion with structures containing similar functionality from across the CSD, including between the 12 ROY polymorphs themselves. However, MGC analysis of the 12 unique ROY polymorphs reveals that the MGC dataset produced for forms **ON** (QAXMEH), **Y** (QAXMEH01), **R** (QAXMEH02), **O** (QAXMEH03), **YN** (QAXMEH04), **ORP** (QAXMEH05) and **YT04** (QAXMEH12) (group 1, 11 structures) is significantly smaller than the same dataset produced for the remaining forms (group 2, 114 structures). Analysis of the structures of the ROY forms in group 1 and group 2 reveals the reason for the discrepancy lies in the representation of the aminothiophene ring; structures in group 1 feature an aromatised aminothiophene ring, whereas structures in group 2 feature discrete (aka kekulised) C=C double bonds. By converting the bonds in group 1 to their kekulised representation, MGC generates the same, larger dataset for all ROY forms. These findings are summarised in Figure 39.



ON
QAXMEH (Aromatised)
Mogul Hits: 12

ON
QAXMEH (Kekulised)
Mogul Hits: 114

Figure 39; (left) simulated structure of the QAXMEH (ON) polymorph of ROY loaded into Mercury with an aromatised thiophene ring, with the number of hits MGC found within the CSD displaying similar geometry. (Right) the same structure loaded with a kekulised thiophene ring, resulting in more similar hits in MGC

Combining the MGC datasets produced for the aromatised and kekulised representations of the ROY forms produces the MGC dataset describing the θ_{thio} values of structurally-related compounds from within the CSD (**MGC Set**). A histogram describing the distribution of θ_{thio} values within the **MGC Set** can be found in Figure 40. It should be noted that the seven structures displaying aromatised structures are all ROY forms: **ON** (QAXMEH), **Y** (QAXMEH01), **R** (QAXMEH02), **O** (QAXMEH03), **YN** (QAXMEH04), **ORP** (QAXMEH05) and **YT04** (QAXMEH12). Table 6 contains a summary of selected crystallographic data for the 12 ROY polymorphs, including CSD refcodes and θ_{thio} angles as calculated using Mogul Geometry Check.

Table 6; Selected crystallographic data, CSD refcodes and θ_{thio} angles for the 12 ROY polymorphs.

Polymorph	CSD Refcode	Z'	θ_{thio} / º	Usual / Unusual Torsion
ON	QAXMEH	1	52.6	Usual
Y	QAXMEH01	1	104.7	Usual
R	QAXMEH02	1	21.7	Usual
O	QAXMEH03	1	46.1	Usual
YN	QAXMEH04	1	104.0	Usual
ORP	QAXMEH05	1	39.4	Usual
YT04	QAXMEH12	1	112.8	Usual
R05	QAXMEH31	2	44.9 (Molecule A) 34.0 (Molecule B)	Usual
PO13	QAXMEH52	1	122.1	Usual
Y04	QAXMEH53	1	76.9	Unusual
R18	QAXMEH57	2	5.0 (Molecule A) 10.3 (Molecule B)	Usual
Y19	QAXMEH60	1	60.7	Usual

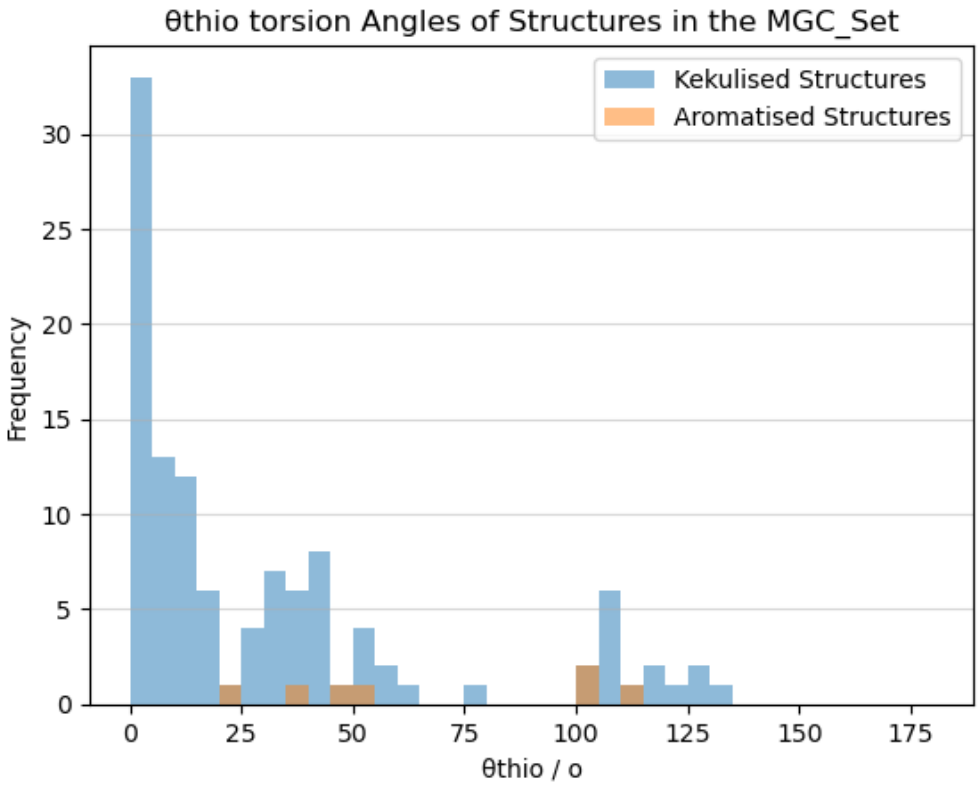


Figure 40; a histogram describing the distribution of inter-ring torsion angles found in a dataset of structures similar to ROY as calculated by Mogul Geometry Check.

However, as the CSD is a repository for all kinds of crystallographic data, the **MGC Set** also contains instances of duplicate polymorphic forms, such as structural redeterminations[81], along with crystals observed at different temperatures and pressures[82]. As this investigation is concerned with unique polymorphic forms, it is important to clean these data in order to provide a more accurate picture of the relevant statistics. Removal of non-unique structures from the **MGC Set** produces a dataset of θ_{thio} values of the unique forms of structurally related compounds to ROY from within the CSD (**Unique Structures Set**) . A histogram describing the distribution of θ_{thio} values within the **Unique Structures Set** can be found in Figure 41.

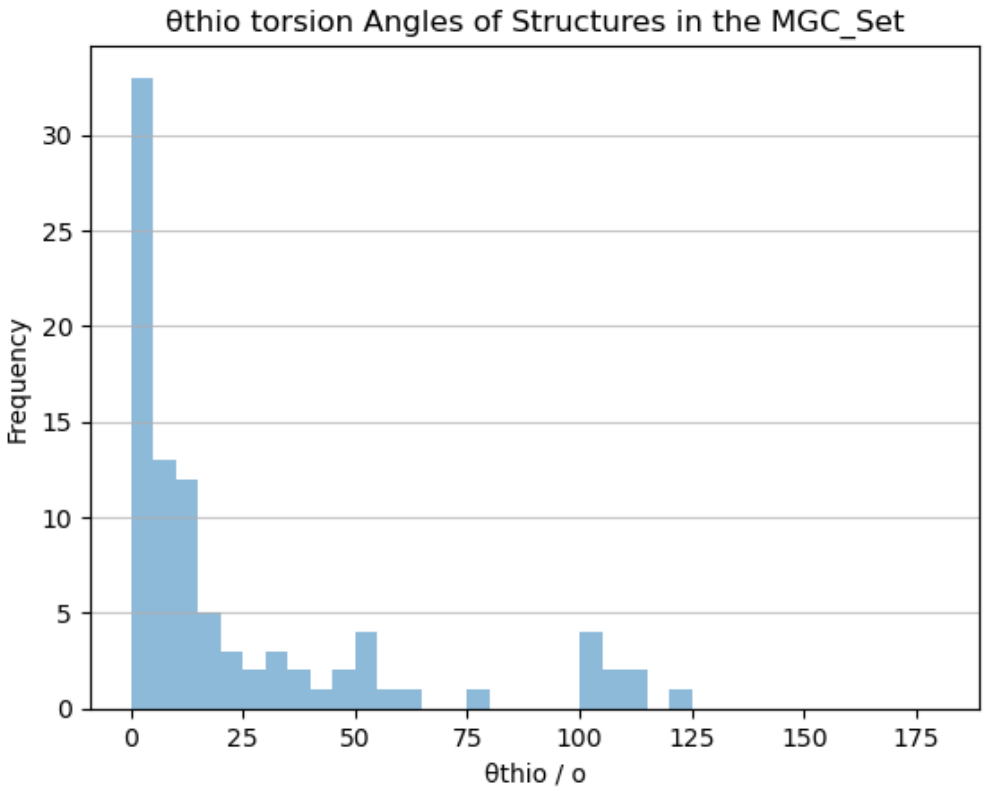


Figure 41; a histogram describing the distribution of θ_{thio} values for unique structures from within the MGS Set calculated for the ROY polymorphs by Mogul Geometry Check

The Unique Structures Set contains 85 structures in a range of torsions, with θ_{thio} values ranging between 0.013° (NIWVOH) to 122.058° (QAXMEH52). There appear to be three clusters of θ_{thio} values, found at $0° < \theta_{thio} < 15°$, $20° < \theta_{thio} < 65°$, and $100° < \theta_{thio} < 125°$. By highlighting the values observed for the 12 unique polymorphs, it becomes apparent that the majority of ROY forms are found in the latter two clusters, with only molecule A and B of the metastable **R18** (QAXMEH57) form of ROY found in the range $0° < \theta_{thio} < 15°$. A histogram of the torsion angles found in the **Unique Structures Set**, along with the 12 unique ROY polymorphs highlighted, can be found in Figure 42.

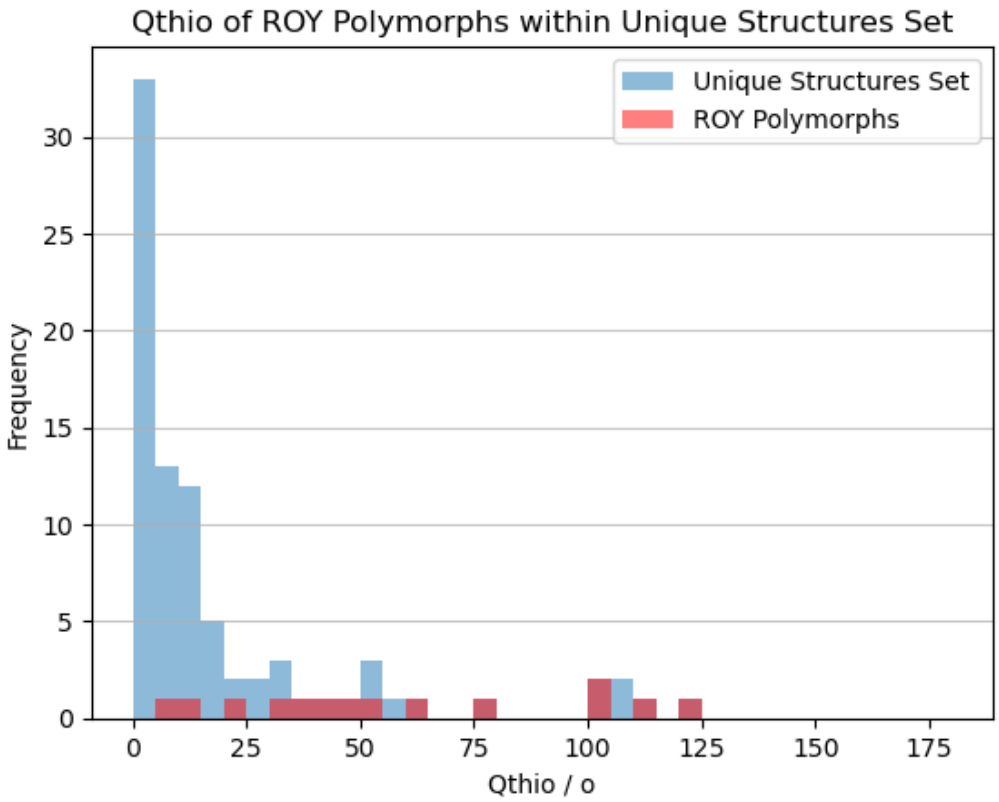


Figure 42; a histogram describing the distribution of θ_{thio} values of structures within the Unique Structures Set, with the θ_{thio} values of the 12 unique ROY polymorphs highlighted in red

It was speculated that structures similar to the ROY scaffold may adopt different θ_{thio} values than those displaying higher differences in their chemical scaffold. In order to test this, the structures within the **Unique Structures Set** were visually analysed in Mercury[83] for their chemical difference to the ROY scaffold. Those structures whose scaffold differed from ROY by a maximum of two heavy atoms were labelled as Matched Molecular Pairs (MMPs) of ROY. **Figure 43** contains a histogram describing the θ_{thio} values of structures within the **Unique Structures Set**, with MMPs of ROY, non-MMPs of ROY and unique ROY polymorphs highlighted.

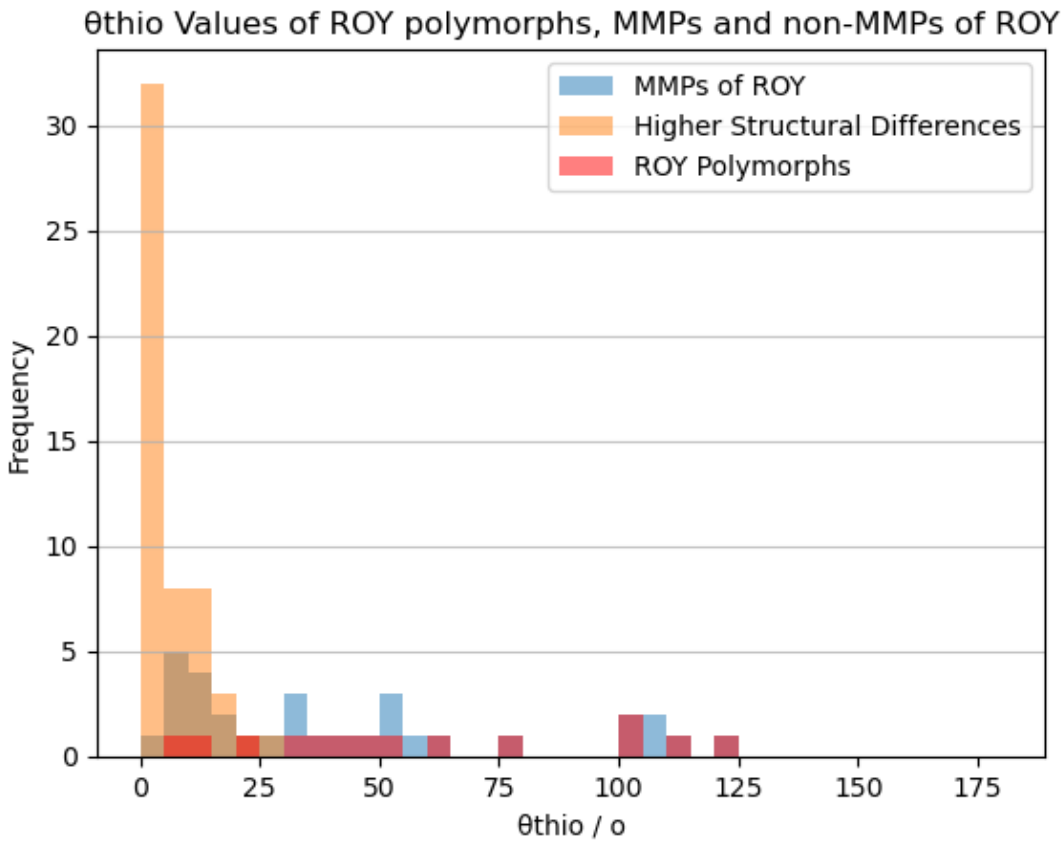


Figure 43; a histogram describing the distribution of θ_{thio} values of structures within the Unique Structures Set, with the θ_{thio} values of the 12 unique ROY polymorphs highlighted in red, MMPs of ROY highlighted in blue and non-MMPs of ROY highlighted in orange.

By highlighting MMPs, non-MMPs and polymorphs of ROY the trends in the observed θ_{thio} values becomes clearer. Non-MMPs of ROY in the **Unique Structures Set** display a distribution of θ_{thio} values which is skewed heavily towards lower θ_{thio} values, or a more planar orientation, with 60% of non-MMP structures (32 of 53 structures) having a θ_{thio} value under 5°, compared with only a single MMP structure (QAXBOK) found in this range. Moreover, all non-MMP structures are found in the range $0° < \theta_{thio} < 30°$, whereas only 41% of MMP structures (13 of 32 structures) and two unique ROY polymorphs found within this range. Furthermore, over 92% of non-MMP structures (49 of 53 structures) are found in the range $0° < \theta_{thio} < 15°$, compared to only 31% of MMPs (10 of 32 structures) and only one unique ROY polymorph, the metastable **R18** (QAXMEH57) form, found in this range.

This analysis would suggest that ROY polymorphs and MMPs of ROY behave differently with respect to the twisting between their constituent aromatic rings to structures displaying higher structural differences (i.e. non-MMPs) of ROY, and that more accurate trend data may be acquired by separating the two

distributions. Figure 44 contains a histogram describing the distribution of θ_{thio} values of unique ROY polymorphs an MMPs of ROY from within the **Unique Structures Set.**

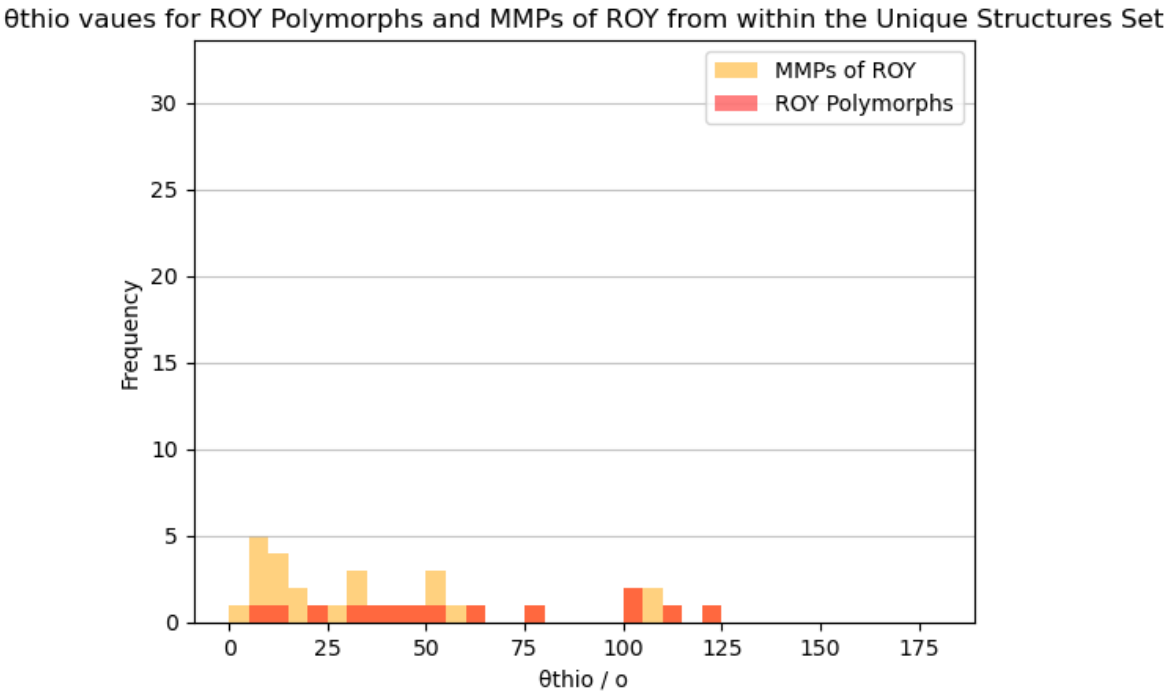


Figure 44; a histogram describing the distribution of θ_{thio} values of MMP-structures and unique ROY polymorphs from within the Unique Structures Set.

It is difficult to take reliable statistics from these data due to the small sample size, however, there are interesting trends to be observed by comparison of ROY polymorphs to MMPs of ROY. The range of θ_{thio} values observed for ROY and MMPs of ROY covers a wider range than is observed for the θ_{thio} values of non-MMPs of ROY, with a distribution of θ_{thio} values which seemingly centre around ~ 35° and 105°. This distribution highlights one structure that falls significantly outside of these distributions, the metastable **Y04** (QAXMEH53) with a θ_{thio} value of 76.87°. This structure was highlighted as displaying an "unusual" torsion via the initial MGC analysis. There also remains a single ROY polymorph which exists just outside of the cluster centred around 105°, belonging to the metastable **PO13** (QAXMEH52) form, which displays a θ_{thio} values of 122.06°. If the MGC function in Mercury was able to automatically clean duplicate structures for its datasets, it is possible that this metastable polymorph would be identified as displaying an "unusual" torsion.

2.3.1 Conformational Potential Energy Surface of ROY

All crystal forms represent local minima on their crystal potential energy surface (PES). Any change in torsion angle results in a new conformation and a new minimum on the crystal PES. Full calculation of the crystal PES is computationally unfeasible, however, for the majority of neutral molecules, equilibrium geometries of molecules within their crystal lattice are close to their gas-phase equilibrium geometries[84]. Moreover, gas-PES conformers are used at starting geometries in the successful prediction of crystal structures of flexible molecules, which has recently been demonstrated for ROY by Beran et al[85].This means that the gas-phase PES can be a useful tool for the evaluation of the conformational PES in the crystalline phase, which can be used to evaluate whether two molecular geometries observed in different crystals correspond to a similar gas-phase conformer.

The gas-PES of ROY was calculated by Cruz Cabeza et al.[86], in a 2014 study into conformational polymorphism. By plotting the θ_{thio} values observed within ROY polymorphs onto the gas-PES it was established that **ON** (QAXMEH), R (QAXMEH), **O** (QAXMEH03) and **ORP** (QAXMEH05) all share a gas phase conformer and are related by a conformational adjustment. The YN, Y and YT04 forms all correspond to a different gas-phase conformer of ROY than ON, and are related to the ON via a conformational change. Figure 45 contains the gas-PES of ROY with the θ_{thio} values of all 12 ROY polymorphs plotted.

Through plotting these torsions on the PES, it becomes apparent that the **R05** (QAXMEH31), **R18** (QAXMEH57) and **Y19** (QAXMEH60) forms of ROY are all related to ON by a conformational adjustment, whilst the PO13 form is a conformational polymorph of **ON**. However, the position of the Y04 (QAXMEH53) form is intriguing . The Y04 form occupies the highest position on the PES relative to all other ROY polymorphs, and is closer to a maximum on the PES than to any of the local minima. This information, coupled with the unusual torsion as highlighted by MGC, could serve as a significant red flag for the stability of the Y04 form.

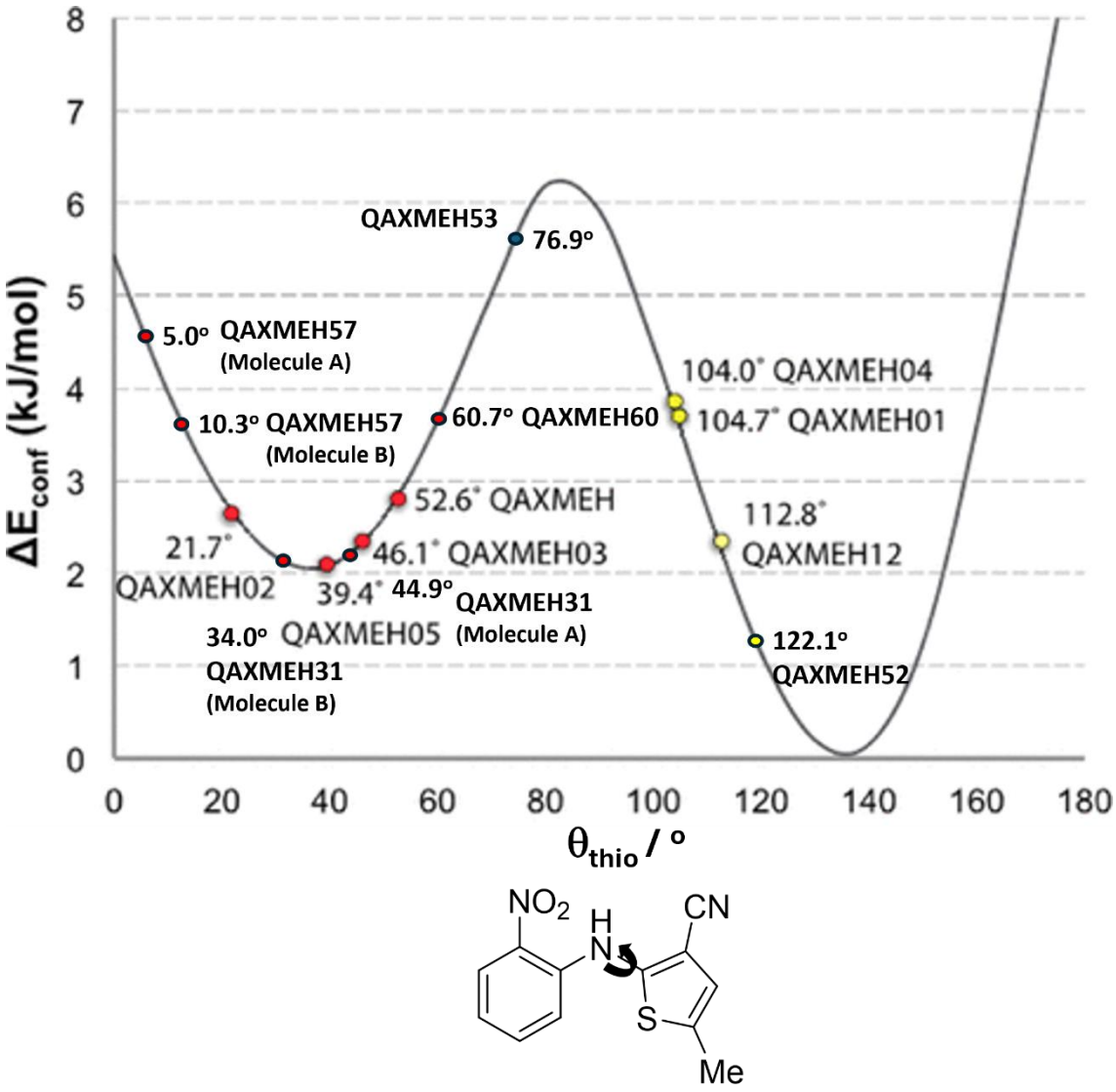


Figure 45; gas-PES of ROY with corresponding θ_{thio} values plotted. Experimental conformations of the ROY polymorphs are plotted as red and yellow circles. Polymorphs QAXMEH – QAMXEH05 were calculated and plotted during the original 2014 study by Cruz-Cabeza et al [86]. Polymorphs QAXMEH31 (molecule A and B), QAXMEH52, QAXMEH53, QAXMEH 57 (molecule A and B) and QAXMEH60 were calculated during this work and added to the original plot. The torsion flagged as unusual by MGC, QAXMEH53, is highlighted as a blue circle. Reprinted (adapted) with permission from [86] Copyright 2014 American Chemical Society.

2.3.2 Intramolecular Environment: Conclusions

Mogul Geometry Check[80] (MGC) is a rapid, effective and easy-to-use informatics tool which allows for the evaluation of various intramolecular factors (i.e. bond length, valence angle, torsion angle and ring geometry) in the context of the Cambridge Structural Database[87] (CSD). In this section MGC was used to analyse the torsion between the two aromatic rings (θ_{thio}) present in the 12 unique ROY polymorphs within the CSD, within the context of structurally-related compounds from the CSD, and MGC was able to rapidly identify the metastable

Y04 (QAXMEH52) form of ROY as displaying an unusual θ_{thio} value. Moreover, analysis of the distribution of θ_{thio} values from the MGC dataset revealed useful information regarding the behaviour of the θ_{thio} value in structures which are matched molecular pairs (MMPs) of ROY vs those with higher structural differences.

However, there are a number of issues with MGC which relate to its use in the solid form informatics (SFI) workflow and its ability to explore structure-polymorphism relationships (SPRs). Firstly, MGC utilises the input structure provided in order to search the CSD for similar structures to use as its basis for comparison. However, as is the case with the ROY polymorphs, there are many equivalent representations which may be used to show the same structure (kekulised vs aromatised aminothiophene ring in ROY, for example), yet these equivalent representations are not present within the MGC search. This necessitates manual alteration of these representations by the user, along with manual combination of the datasets in order to get a more complete picture of similar structures within the CSD. This also increases the likelihood that valuable data are missing from any given analysis with MGC.

The second and final issue relates to the presence of duplicate data within the CSD. The CSD is a public repository for the storage of all kinds of crystallographic data, and as such duplicate data (i.e. structural redeterminations) play an important role in this function. However, when using MGC to investigate SPRs it is important to only include unique crystal forms as to avoid bias in the statistics produced by MGC. Removal of the duplicate structures from within the MGC dataset not only allowed for clearer visualisation of the trends observed between MMPs and non-MMPs of ROY, but also highlights the metastable **PO13** (QAXMEH53) form of ROY as having an unusual θ_{thio} value.

Finally, the use of MGC in tandem with calculations of gas-PES of ROY was able to reaffirm the position of the **Y04** as potentially problematic. This could suggest that the use of traditional computational approaches alongside the informatics-driven MGC could bolster the conclusions which can be drawn from MGC analysis.

2.4 Intermolecular Environment

The intramolecular geometry adopted by a molecule within a crystal structure determines which intermolecular interactions may be formed within that structure, and it is the cumulative effect these intermolecular interactions have across the whole crystal which constitutes the enthalpic contribution to the crystal energy, and hence plays a key role in determining overall polymorphic stability. As such,

during a polymorph screen, analysis of the intermolecular environment of molecules within a crystal structure can provide useful insight into the stability of a given crystal, and hence could be used to supplement the decision-making process when searching for forms with specific properties. Specifically, if molecules within a crystal are found to display only weak interactions, or unusual interactions for the functional groups present, this could be indicative of a metastable form and that further screening may be necessary to find a more thermodynamically stable form.

In this section, we use Mercury[88] to analyse the intermolecular environments of the 12 unique ROY polymorphs contained within the Cambridge Structural Database (CSD)[87]. The visualisation tools within Mercury are used to inspect the intermolecular contacts observed within the crystal structures for each of the ROY polymorphs. Qualitative analysis of these interactions is used to evaluate the intermolecular environment present in each crystal structure relative to the other polymorphs in order to rationalise the thermodynamic trends observed. Once this analysis is complete, the CSD will be utilised to compare the observed interactions for each polymorph to the trends observed across the CSD for similar functional groups using the Full Interaction Maps[89] (FIMs) feature of Mercury. FIMs is used to produce maps detailing the statistical likelihood of finding a certain probe group at a given distance and angle away from a given molecule. Coupling these data with information about the type and directionality of intermolecular interactions observed within each ROY polymorph could provide a more complete picture of the intermolecular environment of a given polymorph, and hence be used to rationalise the thermodynamic trends observed.

2.4.1 Full Interaction Maps and Intermolecular Contacts

By investigating the intermolecular short contacts present between molecules in each of the ROY polymorphs we are able to gain some insight into the intermolecular environment present in each form. This analysis can be used to identify unusual interactions present between molecules, and hence can be of use in rationalising differences in polymorphic stability. There are a number of factors which may determine whether a given intermolecular contact is “favourable”; including differences in electron density, steric crowding around the interacting partners, and the number of interactions each partner is involved in.

In Section 1.2 it was demonstrated that Full Interaction Maps (FIMs) could be used to provide context for the evaluation of the intermolecular contacts observed within the crystal structure of lamotrigine, and hence an evaluation of its overall intermolecular environment. FIMs are an informatics tool constructed from the IsoStar libraries derived from the CSD, in which the distance, angle and

frequency data of intermolecular contacts observed between a central group (molecular fragment) and a given probe contact group (functional group) are plotted. These data provide a 3D contour/heat plot describing the distance, nature and statistical likelihood of finding a given probe group in relation to the input structure. This can serve as an ideal model to describe which intermolecular contacts are expected to form within the crystal lattice of a given input structure, and where they are likely to be found in relation to that input structure. By comparing this idealised statistical model to the intermolecular contacts observed within the crystal lattice of the input structure, it is possible to provide information about how favourable the intermolecular environment present within the crystal lattice of the input structure is.

On its own, this information would be insufficient to make an assessment of polymorphic stability, due to the numerous factors which influence this. However, when multiple polymorphs of the same structure are available, the combination of FIMs and observed intermolecular contacts has potential to assist in the rationalisation of relative polymorphic stability by comparing how far the observed intermolecular contacts differ from the idealised model provided by the FIMs hotspots. To demonstrate this, six ROY polymorphs have been selected for comparison. Three of the most stable ROY forms, **ON** (QAXMEH), **Y** (QAXMEH01) and **O** (QAXMEH03) and three of the least stable forms, **Y04** (QAZMEH53), **R18** (QAXMEH57) and **Y19** (QAXMEH60) have been analysed using FIMs and their respective intermolecular contacts. The probe molecules used to produce the FIM hotspots in this section are uncharged NH hydrogen (blue), carbonyl oxygen (red) and aromatic CH carbon (brown) to represent a hydrogen bond donor, a hydrogen bond acceptor, and a hydrophobic group respectively.

2.4.1.1 FIMs and Intermolecular Contacts: ON, Y and O Forms of ROY

Starting with the stable forms, Figure 46 contains FIMs calculated for the **ON** (QAXMEH) form of ROY, along with the observed intermolecular contacts. **ON** (QAXMEH) is one of the most stable forms of ROY[69] according to available thermal data, a fact which seems to be supported by the location of the observed intermolecular contacts within their associated FIMs hotspots. The dark blue (uncharged NH hydrogen) hotspots associated with the nitrile and nitro groups in **ON** (QAXMEH) are occupied by contacts with two different aromatic protons in favourable interactions. Moreover, the location of these contacts within the red (carbonyl oxygen) hotspots associated with the aromatic protons suggests the reaction is also favourable for the aromatic protons. **ON** (QAXMEH) does not

display any π-type interactions, and hence the brown (aromatic carbon) hotspots remain unoccupied.

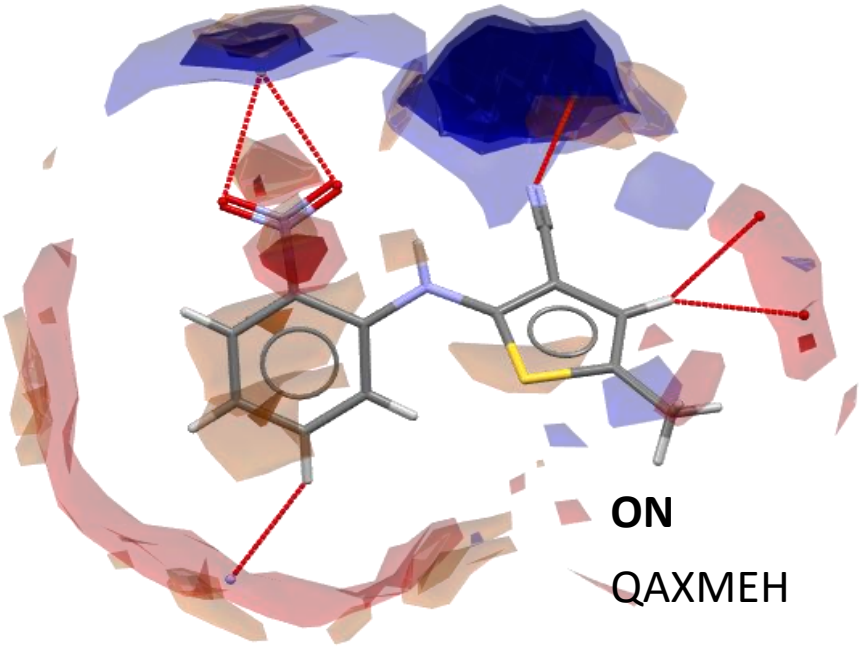


Figure 46; Red (carbonyl oxygen), blue (uncharged NH hydrogen) and brown (aromatic carbon) FIMs hotspots calculated for the ON (QAXMEH) form of ROY. Intermolecular contacts are highlighted by dashed red lines.

FIMs and short contacts calculated for the **Y** (QAXMEH01) form of ROY can be found in Figure 45.

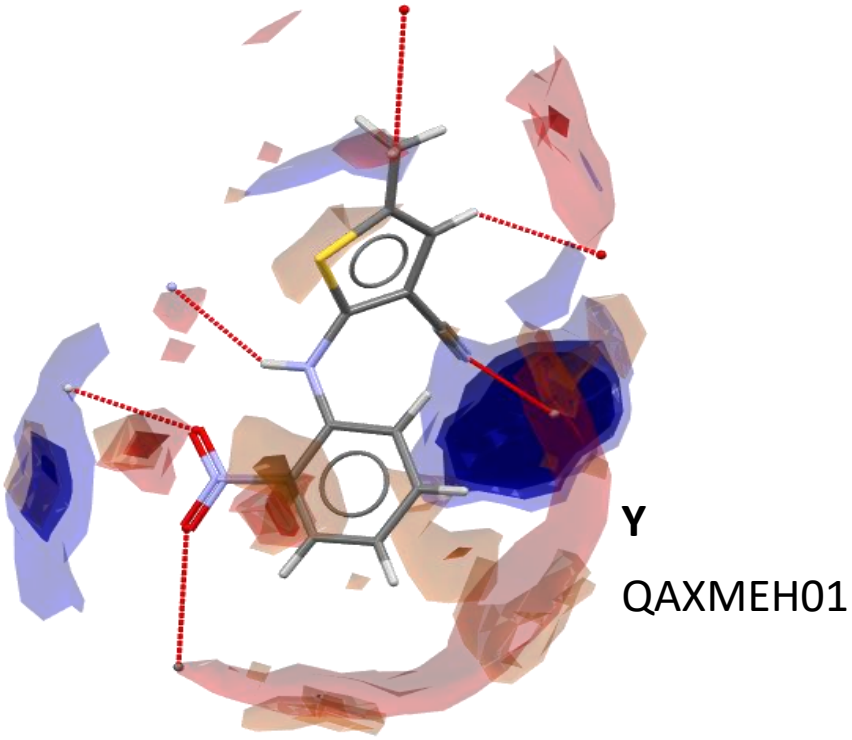


Figure 47; Red (carbonyl oxygen), blue (uncharged NH hydrogen) and brown (aromatic carbon) FIMs hotspots calculated for the Y (QAXMEH01) form of ROY. Intermolecular contacts are highlighted by dashed red lines.

Whilst the overall shape of the hotspots generated for **Y** (QAXMEH01) are similar to those observed for ON (QAXMEH), there are some key differences which may help explain the enantiotropic relationship between the **Y** (QAXMEH01) and **ON** (QAXMEH) forms of ROY.

The major difference between the two is observed between the environments associated with the nitro group and aromatic protons. In **Y** (QAXMEH01), one contact is located at the edge of a red (carbonyl oxygen) hotspot, whilst the other is located at the edge of a blue (uncharged NH hydrogen) hotspot; as opposed to in **ON** (QAXMEH), in which both contacts are found at the centre of a dark blue (uncharged NH hydrogen) hotspot.

The environment of the nitrile group is similar in **ON** (QAXMEH) and **Y** (QAXMEH01), with a single contact located in the centre of the dark blue (uncharged NH nitrogen) hotspot. However, investigation of the atoms involved in these contacts reveals different interaction partners. In **ON** (QAXMEH) the nitrile nitrogen is involved in a contacts with the proton at the 5' position of the nitrobenzene ring, whilst in **Y** (QAXMEH01) the nitrile nitrogen is involved in a contact with the bridging amine proton. A summary of the nitrile and interacting proton environments for **ON** (QAXMEH) and **Y** (QAXMEH01) can be found in Figure 48 and Figure 49, respectively.

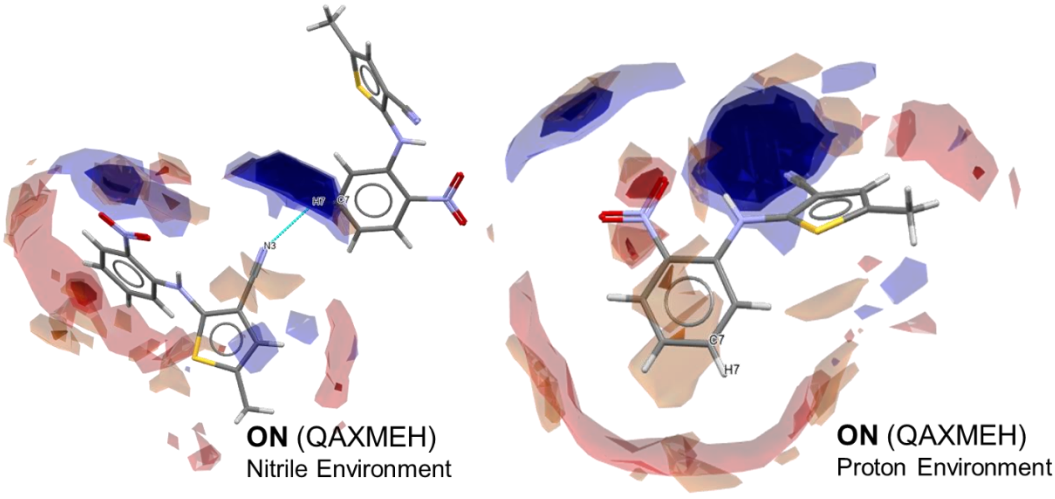


Figure 48; (Left) FIMs and intermolecular contacts calculated for the ON (QAXMEH) Form of ROY with the interaction between the nitrile nitrogen and the proton at the 5' position of the nitrobenzene ring highlighted. (Right) An alternate view of the FIMs calculated for the ON (QAXMEH) Form of ROY, highlighting the red (carbonyl oxygen) hotspot associated with the proton at the 5' position of the nitrobenzene ring.

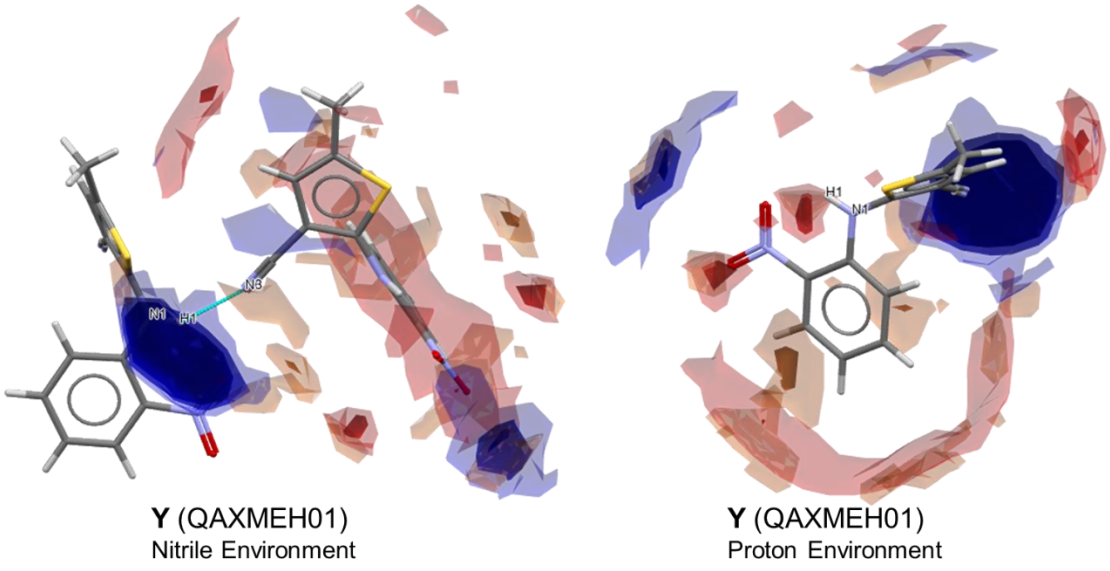


Figure 49; (Left) FIMs and intermolecular contacts calculated for the Y (QAXMEH01) Form of ROY with the interaction between the nitrile nitrogen and the bridging amine proton highlighted. (Right) An alternate view of the FIMs calculated for the Y (QAXMEH01) Form of ROY, highlighting the red (carbonyl oxygen) hotspot associated with the bridging amine proton.

The shade and size of the red (carbonyl oxygen) hotspots associated with the interaction partners of the nitrile nitrogen in **ON** (QAXMEH) and **Y** (QAXMEH01) can provide insight into their favourability. The hotspot associated with the proton at the 5' position of the nitrobenzene ring in **ON** (QAXMEH) is larger than the small red hotspot associated with the bridging amine proton in **Y** (QAXMEH01),

which could suggest a more favourable nitrile environment being present in the **ON** (QAXMEH) Form. This is likely due to the intramolecular H-bond involving the nitro oxygen and bridging amie proton which is present across all known ROY Forms reducing the favourability of interactions between the nitrile nitrogen and bridging amine proton.

The final of the stable ROY forms to discuss is the **O** (QAXMEH03) form. FIMs and short contacts calculated for the **O** (QAXMEH03) form of ROY can be found in Figure 50.

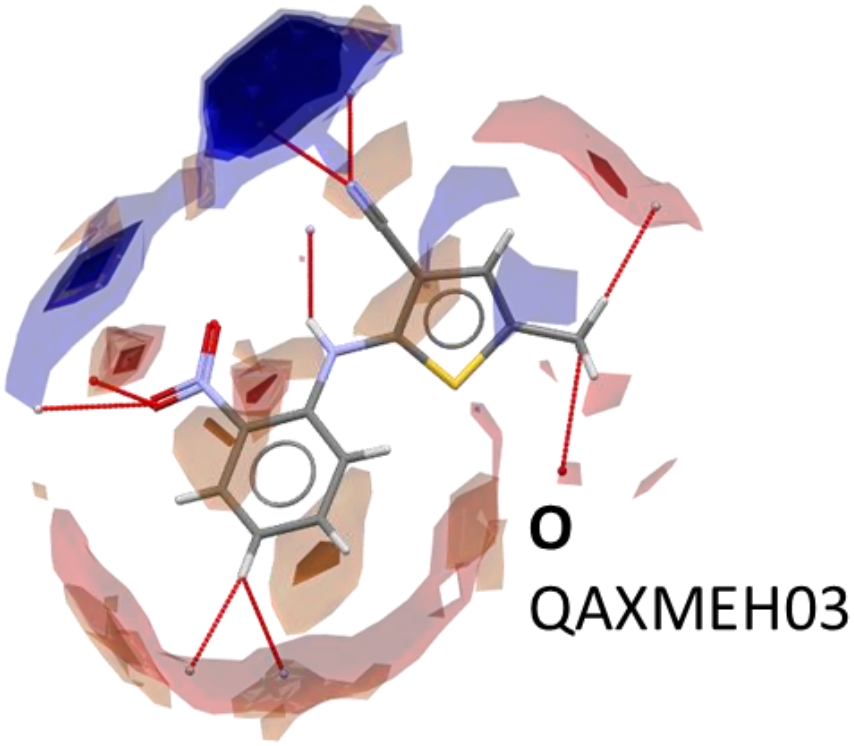


Figure 50; Red (carbonyl oxygen), blue (uncharged NH hydrogen) and brown (aromatic carbon) FIMs hotspots calculated for the O (QAXMEH01) form of ROY. Intermolecular contacts are highlighted by dashed red lines.

The overall environment presented by the FIMs and intermolecular contacts present in **O** (QAXMEH03) contains a mixture of features present across both **ON** (QAXMEH) and **Y** (QAXMEH01) forms. Firstly, there is a contact located in the centre of the dark blue (uncharged NH hydrogen) hotspot, which is observed in all three forms.

The main difference in the intermolecular environment of **O** (QAXMEH03) is the presence of two contacts centred on the nitrile nitrogen. One contact is located at the centre if the dark blue (uncharged NH hydrogen) hotspot, whilst the other is found on the periphery of the same hotspot, resulting in a bifurcated nitrile nitrogen atom. A bifurcated nitrogen atom would theoretically indicate an unfavourable nitrile environment, however, this assessment is incomplete without

first investigating the nature of the atoms/groups involved in the contacts in the wider context of the crystal of the **O** (QAXMEH03) of ROY.

Whilst it is difficult to make an assessment of the relative polymorphic stability of the three stable forms chosen, likely due in part to their enantiotropic thermal relationship, the power of FIMs and intermolecular contacts becomes more apparent when evaluating the intermolecular environments observed for the three metastable ROY forms chosen for this section: **Y04** (QAXMEH52), **R18** (QAXMEH57) and **Y19** (QAXMEH60).

2.4.1.2 FIMs and Intermolecular Contacts: Y04, R18 and Y19 Forms of ROY

The **Y04** (QAXMEH53), **R18** (QAXMEH57) and **Y19** (QAXMEH60) forms of ROY are the most recent forms to be fully characterised; taking the number of unique, single component crystal structures of ROY to a record-holding 12. The **Y04** form was first discovered in 2005 by Chen et al[72] alongside **YT04** (QAXMEH12), however rapid conversion of **Y04** (QAXMEH53) to **YT04** (QAXMEH12) on standing meant that full characterisation remained elusive until 2020, when Li et al[25] utilised encapsulated melt microdroplets to isolate and characterise **Y04** (QAXMEH53).

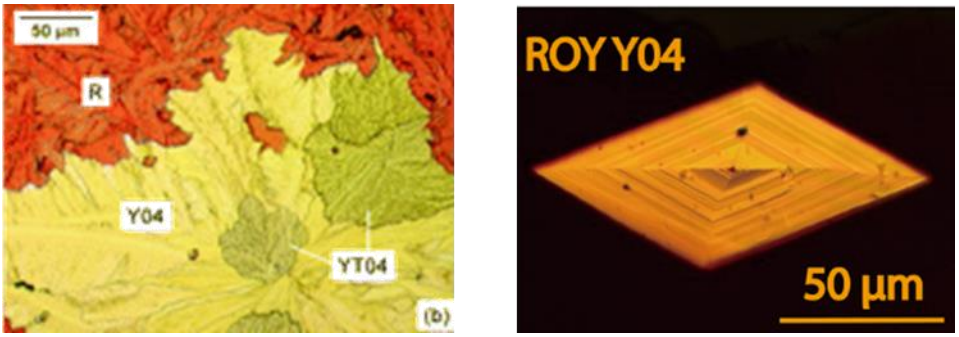
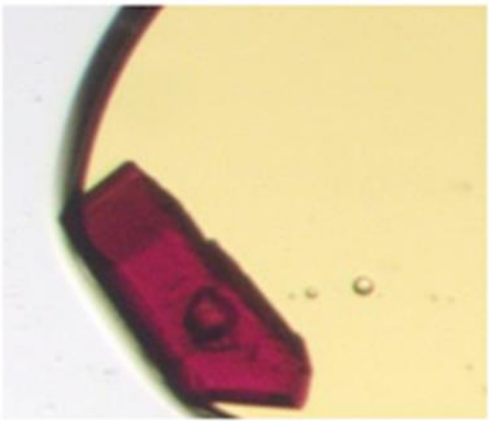


Figure 51; (Left) Concurrent mixtures of R (QAXMEH02), YT04 (QAXMEH12) and Y04 (QAXMEH57), taken from [72]. (Right) Isolated single crystal of Y04 (QAXMEH53), isolated via encapsulated melt microdroplet crystallisation [75].

Also in 2020, a novel crystallisation and screening method produced by Tyler et al, known as encapsulated nanodroplet crystallisation (ENaCt), was used to isolate and characterise the **R18** (QAXMEH57) form of ROY, which was found to be only the second ROY structure for which Z' > 1[75].



R18
QAXMEH57
θ_{thio} = 10.276° and 5.009°
Z' = 2

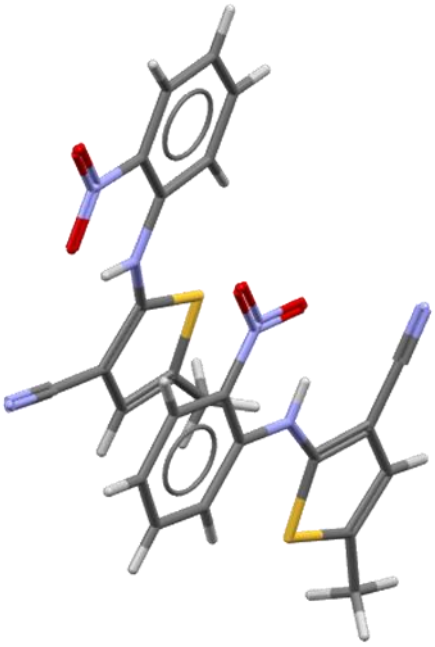
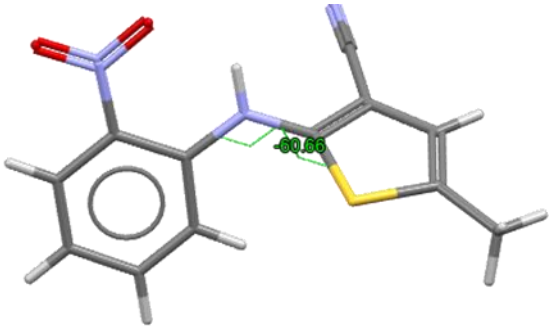
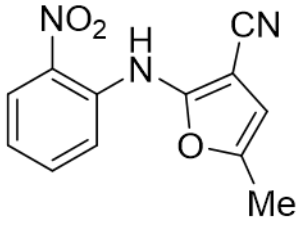


Figure 52; (Left) Isolated crystal of ROY form R18 (QAXMEH57), isolated via ENaCt[75]. (Right) R18 (QAXMEH57) loaded into Mercury, highlighting the two molecules present in the asymmetric unit.

Finally, the **Y19** (QAXMEH60) form of ROY was discovered by Levesque et al, utilising another novel crystallisation method. Levesque et al utilised mixed crystal seeding with crystals of the structurally-related compound Fu-ROY[24]. This novel crystal growth/screening method was used to generate and characterise the **PO13** (QAXMEH52) form, another elusive and metastable form, along with the undiscovered **Y19** (QAXMEH60) form of ROY.



Y19
QAXMEH60
θ_{thio} = 60.66°



FuROY

Figure 53; (Left) Structure of the Y19 (QAXMEH60) form of ROY loaded into Mercury. (Right) Chemical structure of the structurally-related compound FuROY, which was utilised in mixed crystal seeding experiments in order to generate Y19 (QAXMEH60).

The elusive nature of these three forms, along with the niche methods used to access them, highlights the metastability of the **Y04** (QAXMEH53), **R18**

(QAXMEH57) and **Y19** (QAXMEH60) forms relative to the easily isolated and characterised **ON** (QAXMEH), **Y** (QAXMEH01) and **O** (QAXMEH03) forms of ROY. Moreover, through analysis of the FIMS and intermolecular contacts for each of the metastable forms, key differences in the intermolecular environments of the stable and metastable forms of ROY become apparent.

The first of the metastable forms to discuss is the **Y04** (QAXMEH53) form of ROY, for which FIMs and intermolecular contacts have been highlighted in Figure 54.

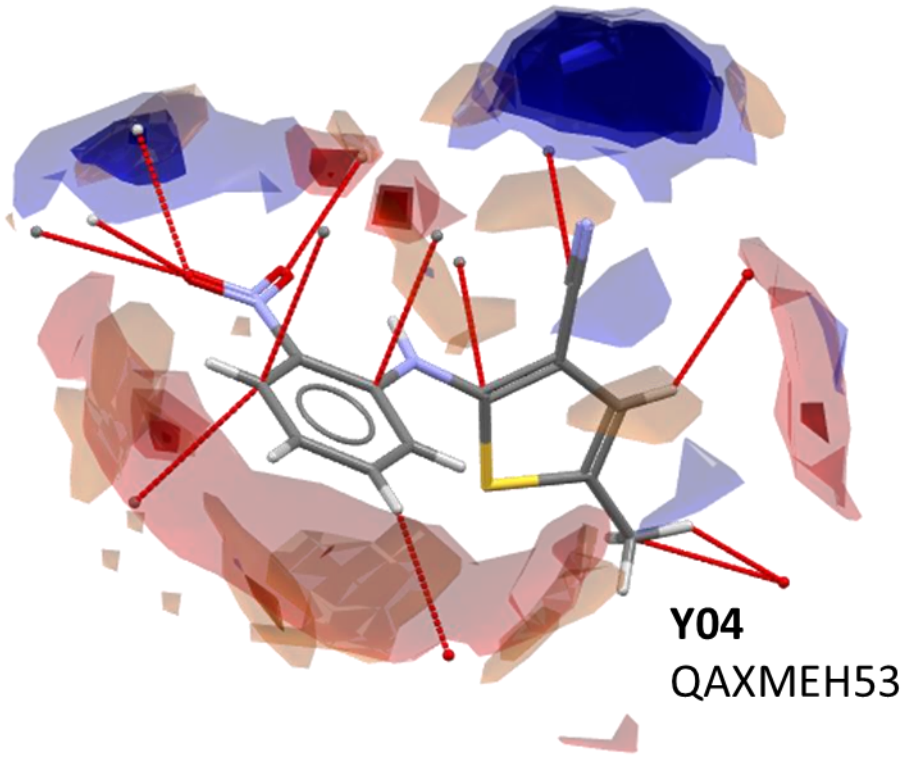


Figure 54; Red (carbonyl oxygen), blue (uncharged NH hydrogen) and brown (aromatic carbon) FIMs hotspots calculated for the Y04 (QAXMEH53) form of ROY. Intermolecular contacts are highlighted by dashed red lines

At a glance, there are numerous differences between the intermolecular environment presented by **Y04** (QAXMEH53) compared to the (QAXMEH), **Y** (QAXMEH01) and **O** (QAXMEH03) ROY forms (herein referred to as stable forms). The most immediately obvious is the high proportion of intermolecular contacts located outside of their nearest FIMs hotpots, suggesting the presence of statistically less likely contacts formed in **Y04** (QAXMEH53) compared to the stable forms of ROY. The next difference is observed around the nitro and nitrile groups. The nitro group in **Y04** (QAXMEH53) is involved in four discrete intermolecular contacts, double the highest number of contacts observed for any stable form. Conversely, the nitrile nitrogen atom forms zero intermolecular contacts, leaving the entire FIMs hotspot unoccupied in **Y04** (QAXMEH53). Finally, there appear to be contacts formed above/below the plane of the benzene

ring of **Y04** (QAXMEH53), another feature not observed across any of the stable ROY forms.

FIMs hotspots and intermolecular contacts for **R18** (QAXMEH57) can be found in Figure 55. **R18** (QAXMEH57) is one of only two ROY forms for which Z' > 1, with two molecules in its asymmetric unit. Both molecules in the asymmetric unit of **R18** (QAXMEH57) are found to display identical intermolecular contacts, except for the nitrile group in molecule B, and hence only the intermolecular environment of molecule B of **R18** (QAXMEH57) is discussed.

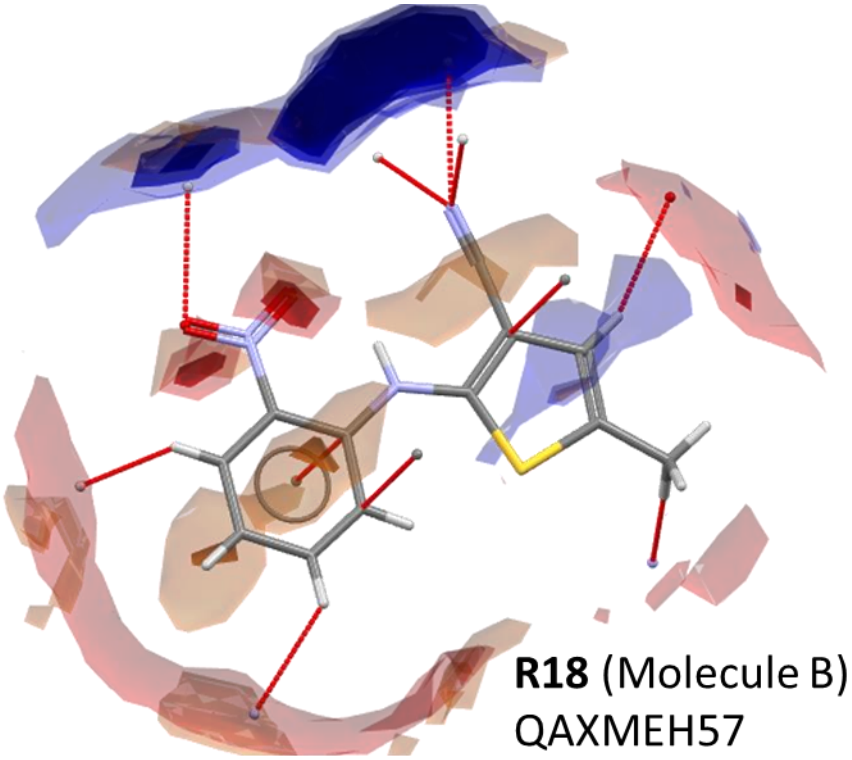


Figure 55; Red (carbonyl oxygen), blue (uncharged NH hydrogen) and brown (aromatic carbon) FIMs hotspots calculated for molecule B of the R18 (QAXMEH57) form of ROY. Intermolecular contacts are highlighted by dashed red lines

The overall intermolecular environment displayed by **R18** (QAXMEH57) is more similar to the stable forms of ROY than that displayed by **Y04** (QAXMEH53), with a higher proportion of contacts located inside FIMs hotspots in the **R18** (QAXMEH57) form compared to **Y04** (QAXMEH53). However, the more striking difference to the stable forms displayed by **R18** (QAXMEH57) is the three intermolecular contacts centred on the nitrile nitrogen atom. This results in a trifurcated nitrogen atom and would likely lead to an unfavourable interaction environment for **R18** (QAXMEH57).

FIMs hotspots and intermolecular contacts for **Y19** (QAXMEH60) can be found in Figure 54. The overall intermolecular environment present in **Y19** (QAXMEH60) shares a number of similarities with the other metastable forms **Y04** (QAXMEH53) and **R18** (QAXMEH57).

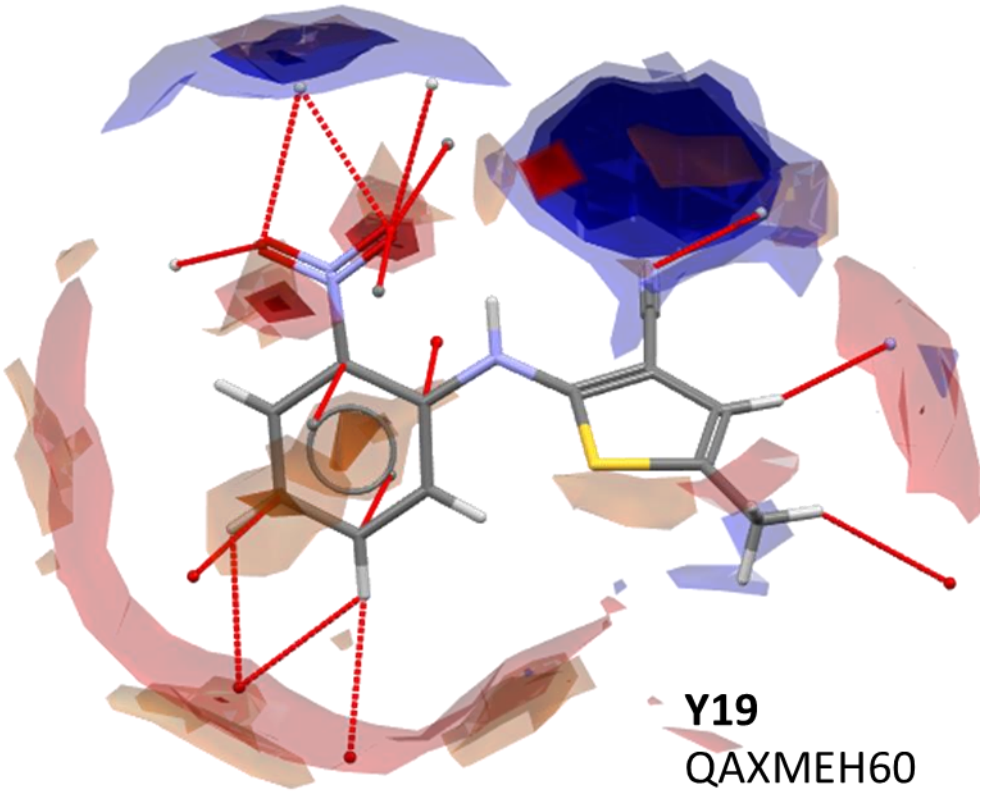


Figure 56; Red (carbonyl oxygen), blue (uncharged NH hydrogen) and brown (aromatic carbon) FIMs hotspots calculated for the Y19 (QAXMEH60) form of ROY. Intermolecular contacts are highlighted by dashed red lines

The first similarity to the metastable forms is observed in the high proportion of intermolecular contacts located outside of their nearest FIMs hotspots, similar to the environment displayed by **Y04** (QAXMEH53), suggesting an overall statistically less-common environment. **Y19** (QAXMEH60) also displays a highly crowded nitro environment, display four discrete intermolecular contacts, similar to the nitro group of **Y04** (QAXMEH53) and the nitrile group of molecule B of **R18** (QAXMEH57). Moreover, similar to the benzene group of **Y04** (QAXMEH53), there are contacts located in the brown (aromatic CH carbon) FIMs hotspot, a feature not present in any of the stable forms discussed in 2.2.1.1.

2.4.1.3 FIMs and Intermolecular Contacts: Conclusions

Analysis of the observed intermolecular contacts formed by different ROY polymorphs in the context of Full Interaction Maps (FIMs) hotspots generated for

each form has revealed a number of intermolecular features which tend to correlate with their relative polymorphic stability. Comparison of the intermolecular environments of the **ON** (QAXMEH), **Y** (QAXMEH01) and **O** (QAXMEH03), or *stable*, forms or ROY, with the **Y04** (QAXMEH53), **R18** (QAXMEH57) and **Y19** (QAXMEH60), or *metastable*, forms, revealed that the intermolecular contacts of the stable forms are primarily located within the FIMs hotspot of their corresponding group, whereas the majority of contacts of metastable forms are located outside of these hotspots. Stable ROY polymorphs form contacts in the most dense regions of the most prominent FIMs hotspots (i.e. around the nitrile and nitro groups), whilst metastable polymorphs form contacts on the periphery of these hotspots, or not at al.

It was found that metastable forms are likely to display significant crowding around prominent groups, such as the nitro groups in **Y04** (QAXMEH53) and **Y19** (QAXMEH60), and the nitrile group of molecule B of **R18** (QAXMEH57), which may be an indicator of an unfavourable intermolecular environment. Conversely, contacts involving the bridging amine group were only present in the stable forms **Y** (QAMEH01) and **O** (QAXMEH03), suggesting the importance of this contact to the overall stability of ROY polymorphs.

Finally, it was noted that interactions located in/near the brown (aromatic CH carbon) hotspots were only observed for the metastable forms **Y04** (QAXMEH53) and **Y19** (QAXMEH60), which could suggest the correlation of π-π stacking-type interactions with the presence of more metastable ROY forms.

Ultimately, these analyses are limited to a small number of the available ROY polymorphs, and further investigation of the intermolecular environment of the remaining ROY polymorphs would be necessary to properly evaluate these trends.

2.5 Supramolecular Environment

The supramolecular environment of a crystal describes the cumulative effect of individual intermolecular-interactions on the properties of the crystal. These properties include many factors which play a role in the processability and developability of a given crystal form, such as thermodynamic stability, mechanical and processing behaviour.

Supramolecular environment is concerned not only with the strength of a given interaction, but also its directionality and overall connectivity within the crystal network. As such, one of the most important intermolecular interactions to consider when modelling the supramolecular environment of a crystal is the hydrogen-bond. H-bonds are not only relatively strong interactions in neutral

molecular systems, but are also highly directional, allowing them to form multidimensional networks throughout a crystal structure. This can have a profound effect on not only the thermodynamic stability of a crystal, but also on its mechanical behaviour and hence its downstream processability. As such, an understanding of the H-bond environment present within a crystal structure can provide useful information regarding which form may be optimal for further development during a pharmaceutical polymorph screen.

In this section, Hydrogen-Bond Propensity (HBP) modelling[90] is used to assess the H-bonding environments present within each of the ROY polymorphs present within the Cambridge Structural Database (CSD)[87]. The goal of these analyses is to provide a description of the supramolecular environment present in each polymorph, and to ascertain whether this description provides sufficient information to contextualise the differences in polymorphic stability observed between different ROY polymorphs. Along with their predictive capabilities, the HBP modelling tools will be assessed for their applicability, interpretability and ease of use.

2.5.1 Hydrogen-Bond Propensity Modelling

The first step in the HBP analysis involves the automatic detection of H-bond donor and acceptor groups from the input structure, which are converted into donor/acceptor fragments which are used to mine the CSD for structures containing similar functionality, and it is here that we encounter our first issue. As was the case with Mogul Geometry Check[80] in section 2.1, the different structural representations found for the different polymorphs (see aromatised vs kekulised) results in slight differences in the resultant HBP models. Namely, the aromatised thiophene ring found in **ON** (QAXMEH), **Y** (QAXMEH01), **R** (QAXMEH02), **O** (QAXMEH03), **YN** (QAXMEH04), **ORP** (QAXMEH05) and **YT04** (QAXMEH12) results in the generation of different fragments for the secondary amine, nitrile and cyclic thioether groups, compared to those generated from forms containing kekulised thiophene rings.

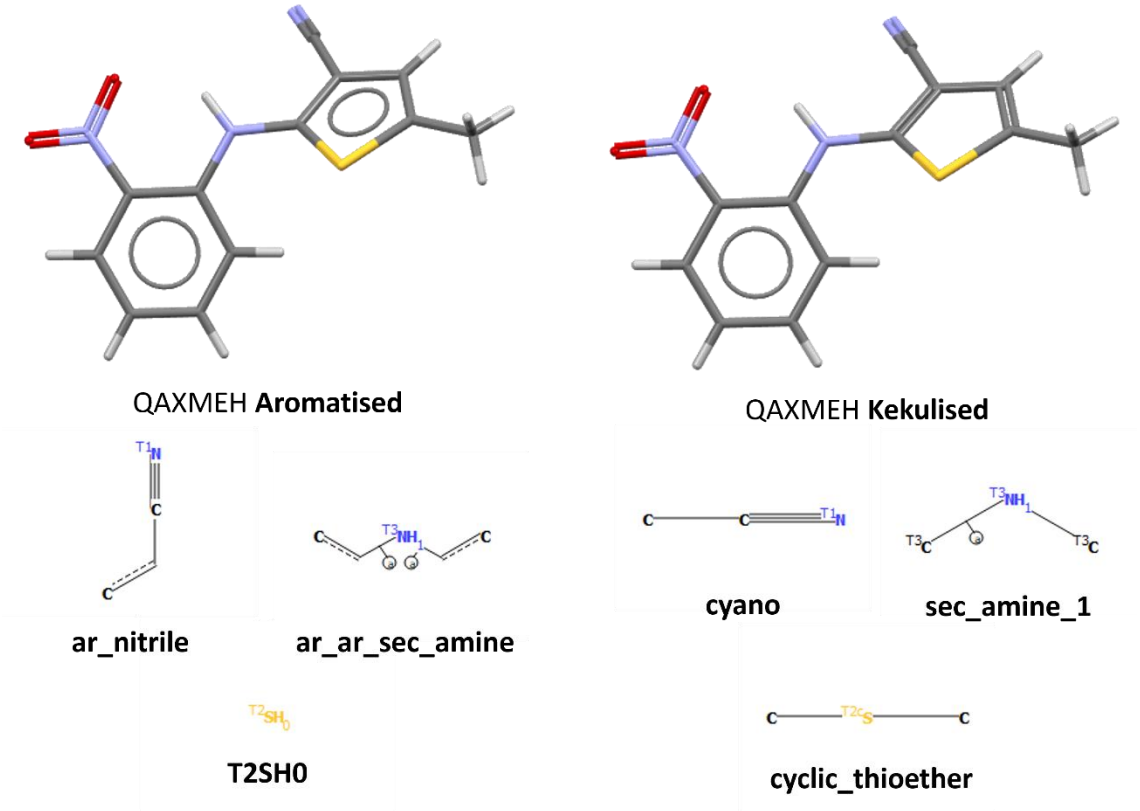


Figure 57; (Left) Aromatised representation of ON structure of ROY, along with the resultant different fragments generated by the HBP model and their corresponding labels. (Right) Kekulised representation of the same ON structure of ROY, along with the different fragments generated by the HBP model and their corresponding labels.

Since the kekulised representation yielded the best results in the Mogul Geometry Check search it is assumed that this is the more commonly observed form of the two within the CSD, and hence all structures were converted to their kekulised forms before proceeding with the HBP. After conversion to the kekulised form all structures produce the same fragments, and, except those cases where Z' = 2 **R05** (QAXMEH31) and **R18** (QAXMEH57) the same number of donor and acceptor groups. Figure 58 contains the donor and acceptor fragments used to produce HBP models for the twelve unique ROY polymorphs within the CSD, along with their assigned labels.

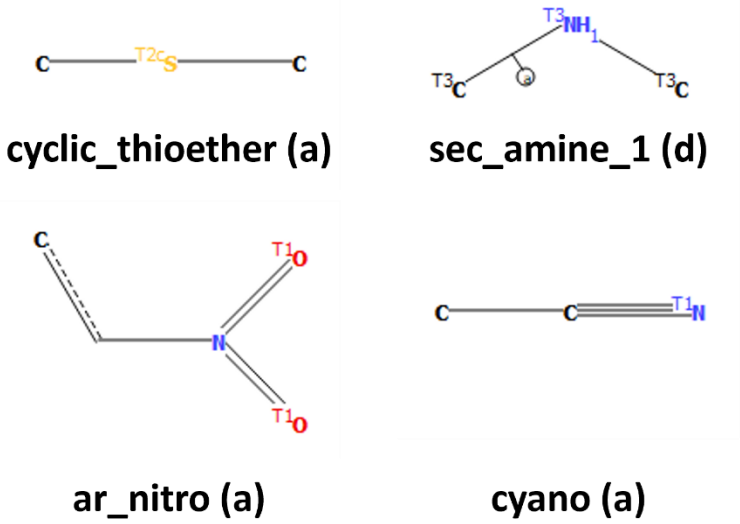


Figure 58; Fragments generated for each ROY structure. These are used to mine the CSD for H-bond statistics along with constructing potential H-bonding networks through combinations of donor and acceptor fragments. Labels are given below each fragment. Donor and acceptor groups are denoted as (d) and (a), respectively.

The HBP modelling software initially attempts to mine the CSD for examples of **sec_amine_1 (a)** but returns zero examples in 5141 structures, likely due to the planarity of the C-N-C bond resulting in an inaccessible nitrogen lone-pair, and hence the **sec_amine_1 (a)** factor is removed from the HBP model. The HBP model uses statistics of structures containing similar functionality to model the predicted propensity towards the formation of an intermolecular H-bond for each donor/acceptor pairing possible within Figure 58, calculating a propensity score for each pairing. This is accompanied by a calculation of the statistical likelihood of forming zero, one or two intermolecular H-bonds for each donor/acceptor atom within the ROY structure. A static model, based on factors such as steric crowding, donor/acceptor ratio and the presence intramolecular H-bonds intramolecular H-bonds, is used to calculate a coordination score for each of the zero, one and two intermolecular H-bonds states for each atom, which serves as a measure of preferred intermolecular connectivity for each donor/acceptor atom within the ROY structure. By taking the average of these two metrics (propensity score and coordination score) we can assess the favourability of the H-bonding network which is predicted by the HBP model, which can be used as an assessment of the supramolecular environment associated with each network, with higher mean propensity and coordination scores being indicative of a more favourable supramolecular environment.

Figure 59 contains a scatterplot of the predicted H-bonding networks constructed by the HBP model along with their associated average propensity and coordination scores. To reduce redundancy and increase the readability of the

plot, only those structures with average coordination scores ≥ 0.2 have been shown. Networks consisting of zero H-bonds (purple circles), NH + CN H-bonds (blue crosses), NH + NO2 H-bonds (green crosses), and networks from structures for which Z' = 2 (grey plusses) are highlighted.

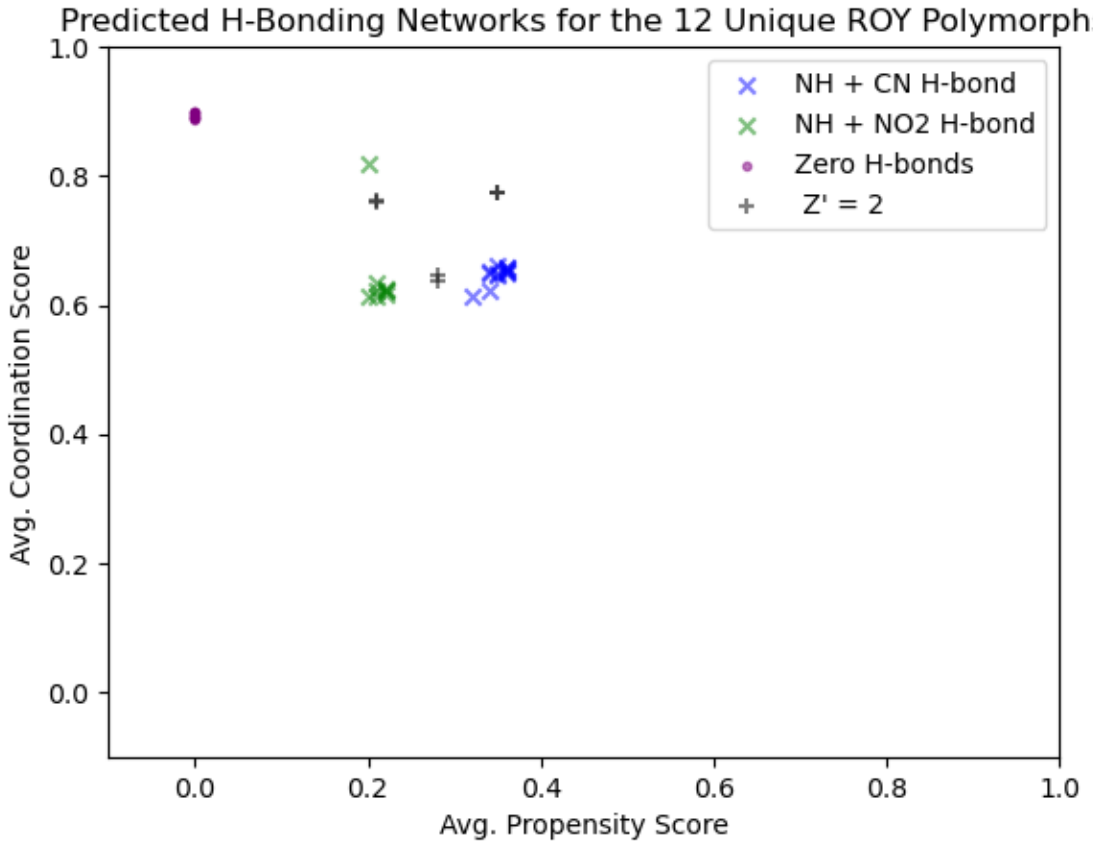


Figure 59; a scatterplot describing the average coordination and propensity scores for the predicted H-bonding networks generated by the Hydrogen Bond Propensity model for the 12 unique ROY polymorphs within the CSD. Networks belonging to structures for which Z' = 2 (yellow), H-bonds between sec_amine_1 (d) + ar_nitro (a) (orange), H-bonds between sec_amine_1 (d) + cyano (a) (red), and containing zero intermolecular H-bonds (black) are highlighted.

There are three main clusters corresponding to the three most likely H-bonding networks to be adopted by the ROY polymorphs. The first cluster is found towards the top left of the graph and is associated with the case where zero intermolecular H-bonds are formed, thus resulting in the average propensity score of zero. This cluster is associated with the highest average coordination scores calculated by the model, suggesting the most favourable connectivity of each of the donor and acceptor atoms within the ROY structure is to form zero intermolecular H-bonds.

The next region to discuss is found towards the right-hand side of the graph and consists of networks made up of isolated **sec_amine_1 (d)** + **cyano (a)** H-bonds. The networks in this region display the highest average propensity score, ranging

from 0.32 to 0.36, which is simply the propensity score of the **sec_amine_1 (d)** + **cyano (a)** H-bond. This score is relatively low, suggesting the **sec_amine_1 (d)** + **cyano (a)** H-bond is not particularly favourable. This makes sense; due to the presence of a 6-membered ring intramolecular H-bond in the ROY structure, any intermolecular H-bond involving the **sec_amine_1 (d)**, the only donor in the structure, would have a bifurcated donor group. The average coordination score for these networks is also smaller than the zero intermolecular H-bond cluster, reflecting the predicted preference for the ROY structure against the formation of intermolecular H-bonds.

The next cluster of networks is found around an average propensity score of 0.21 – 0.24. This cluster corresponds to a network consisting of isolated **sec_amine_1 (d)** + **ar_nitro (a)** H-bonds. The lower average propensity scores for networks found in this cluster suggests that the **sec_amine_1 (d)** + **ar_nitro (a)** H-bond is less favourable than the **sec_amine_1 (d)** + **cyano (a)** H-bond, resulting in a less favourable H-bonding network than those found in the **sec_amine_1 (d)** + **cyano (a)** H-bond cluster. This is another reasonable suggestion, given that the **ar_nitro (a)** atom is also involved in the same 6-membered ring intramolecular H-bond that the **sec_amine_1 (d)** atom is involved in, both donor and acceptor atoms are bifurcated, making this H-bond even less likely than the **sec_amine_1 (d)** + **cyano (a)** H-bond. The average coordination scores in this cluster are roughly the same as those associated with the **sec_amine_1 (d)** + **cyano (a)** H-bond cluster, further highlighting the preference of the ROY structure towards the formation of zero intermolecular H-bonds.

The final aspect of the potential H-bonding network scatterplot to discuss is the few networks which fall outside of the previously mentioned clusters. The first of these is a pair of networks found at an average propensity score of 0.28. These networks correspond to the **R05** (QAXMEH31) and **R18** (QAXMEH57) structures, which are the only Z = 2 polymorphs of ROY available within the CSD. Due to the presence of multiple structures within the asymmetric unit, the HBP model calculates for the situation where both molecules within the crystal lattice are involved in H-bonds with different bonding partners. In this case the networks refer to the H-bonding environment in which one **sec_amine_1 (d)** + **cyano (a)** H-bond and one **sec_amine_1 (d)** + **ar_nitro (a)** H-bond exist within the same structure. As the propensity score for the sec_amine_1 (**d)** + **ar_nitro (a)** H-bond is 0.21 and the propensity score of the **sec_amine_1 (d)** + **cyano (a)** H-bond is 0.35, this results in an average propensity score of 0.28. The **R05** (QAXMEH31) or **R18** (QAXMEH57) structures are also responsible for a further two clusters with anomalously large average coordination scores, located at average propensity scores of 0.22 and 0.35. However, in these cases, deviation from the

general trends is a result of the average coordination score calculation. Due to the presence of two molecules within the asymmetric unit, the coordination model produces coordination scores for ten donor/acceptor groups (**N1 of sec_amine_pl (d)**, **N3 of nitrile (a)**, **O1 of nitro (a)**, **O2 of nitro (a)** and **S1 of cyclic_thioehter (a)** for each molecule). Whilst this has no effect on the coordination scores of individual donor/acceptor groups, it can have an effect on the average coordination score if both ROY molecules within the crystal lattice of **R05** (QAXMEH31) or **R18** (QAXMEH57) adopt different H-bonding environments. The cluster at an average propensity score of 0.21 refers to the case where one molecule in **R05** (QAXMEH31) or **R18** (QAXMEH57) is involved in a **sec_amine_1 (d)** + **ar_nitro (a)** H-bond, whilst another forms zero intermolecular H-bonds. This results in an average coordination score of around 0.75 as opposed to around 0.62, as one molecule within the crystal lattice of **R05** or **R18** would be in a more favourable coordination environment. This situation is similar to the cluster observed around an average propensity score of 0.35, except this case refers to the situation where one molecule in **R05** (QAXMEH31) or **R18** (QAXMEH57) forms a **sec_amine_1 (d)** + **cyano (a)** H-bond, whilst the other forms zero intermolecular H-bonds.

In order to evaluate the ability of HBP modelling to assess the favourability of the H-bonding environments of the ROY polymorphs it is essential to compare the networks predicted by the HBP model to those observed in each ROY polymorph. Figure 60 contains a scatterplot of the average coordination scores and propensity scores for the observed H-bonding networks of the twelve unique ROY polymorphs from within the CSD. The first noticeable thing about the two distributions is the striking similarities between the two. This makes sense, as the low number of potential H-bonding partners results in a small number of potential networks for HBP to predict, whilst the relatively large number of solved structures of ROY increases the likelihood that multiple H-bonding networks will be discovered. Nonetheless, the fact all of the observed H-bonding networks adopted by ROY polymorphs fall into categories predicted by HBP makes ROY a useful case study.

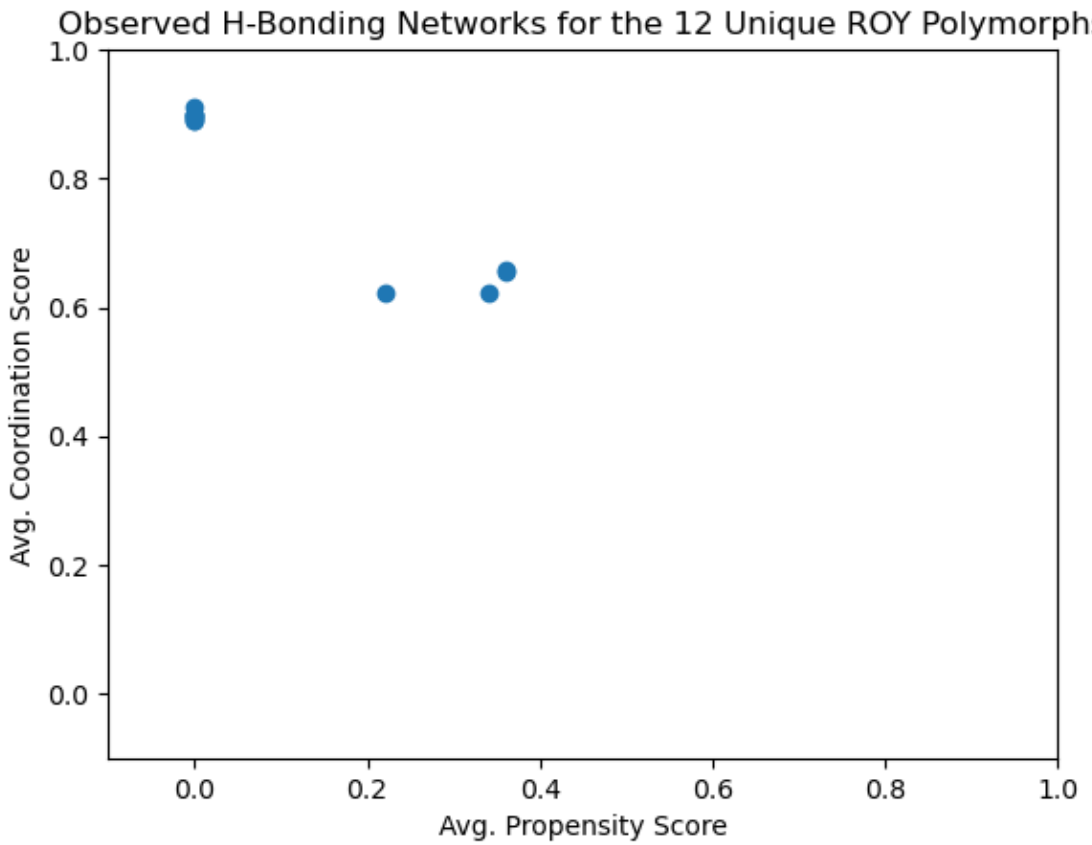


Figure 60; a scatterplot describing the average coordination and propensity scores for the predicted H-bonding networks generated by the Hydrogen Bond Propensity model for the 12 unique ROY polymorphs within the CSD.

The observed H-bonding networks of ROY polymorphs fall into two categories: those which display an intermolecular H-bond (**Y** (QAXMEH01), **O** (QAXMEH03), **YN** (QAXMEH04) and **YT04** (QAXMEH12)), and those which do not (**ON** (QAXMEH), **R** (QAXMEH02), **ORP** (QAXMEH05), **R05** (QAXMEH31), **PO13** (QAXMEH52), **Y04** (QAXMEH53), **R18** (QAXMEH57) and **Y19** (QAXMEH60)). Of the four forms found to display intermolecular H-bonds, these can be further divided by the partners involved in the H-bonding network.

The **YN** (QAXMEH04) form displays an intermolecular H-bond between **sec_amine_1 (d)** and **ar_nitro (a)**, resulting in an average propensity score of 0.21, which can be found in Figure 61. The lower propensity score in this case is likely due to the presence of the six-membered ring intramolecular H-bond in which both O1 and N2 atoms are involved, resulting in an intermolecular H-bond in which both donor and acceptor atoms are bifurcated.

The three remaining forms, **Y** (QAXMEH01), **O** (QAXMEH03) and **YT04** (QAXMEH12), all display the same intermolecular H-bond between **sec_amine_1 (d)** and **cyano (a)**, which can be found with an average propensity score of ~ 0.35. A visual representation of this H-bond can be found in Figure 62.

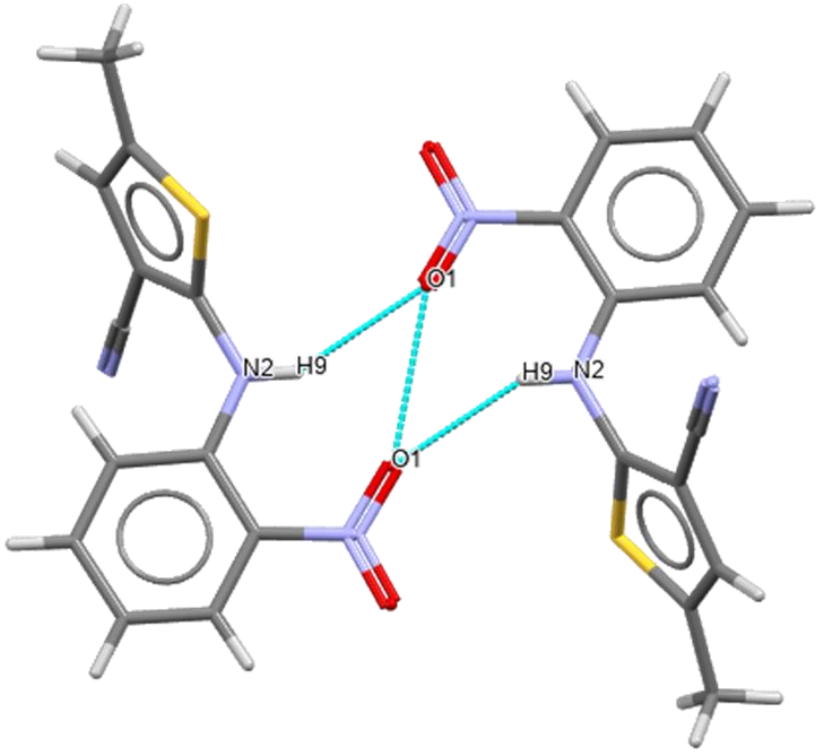


Figure 61; a visualisation of the sec_amine_1 (d) + ar_nitro (a) H-bond observed in YN (QAXMEH04).

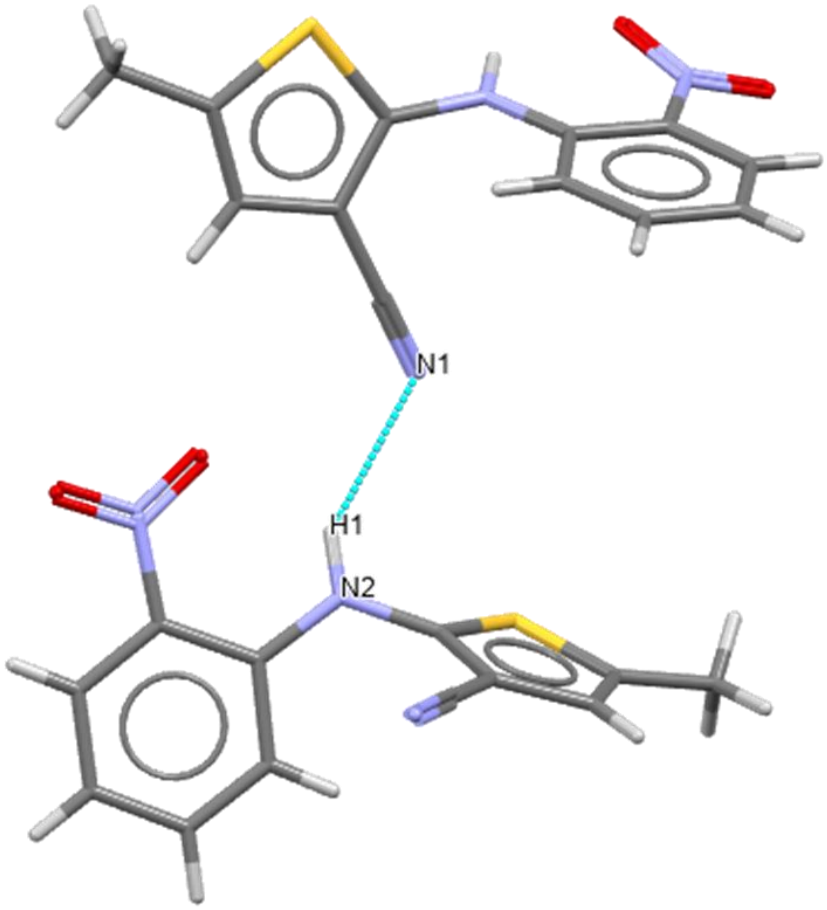


Figure 62; a visualisation of the sec_amine_1 (d) + cyano (a) H-bond observed in YT04 (QAXMEH12)

According to the thermal data collected by Yu et al., **YN** (QAXMEH04) is the least thermally stable form of the six initially discovered ROY forms, whilst the **Y** (QAXMEH01) and **O** (QAXMEH03) forms are two of the most stable forms[69]. The **YT04** (QAXMEH12) form of ROY was not formally discovered and isolated until 2005 by Chen et al., however, it was discovered that the **YT04** (QAXMEH12) form was the most dense form of ROY at 25 °C, along with being the second most stable at 0 K, behind **Y** (QAXMEH01)[70].

These data would suggest that the presence of an intermolecular H-bond between the **sec_amine_1 (d)** and **cyano (a)** groups of ROY correlate with thermodynamically stable forms, due to the relative polymorphic stability of all three forms displaying this intermolecular H-bond. The data do not suggest that the presence of *any* intermolecular H-bond is correlated with stable forms of ROY, highlighted by the relative metastability of the **YN** (QAXMEH04) form, nor does it suggest the absence of intermolecular H-bonds being correlated with lower polymorphic stability, highlighted by the high stability of the **ON** (QAXMEH) form, which itself displays zero intermolecular H-bonds. These data serve to highlight

the important role the **sec_amine_1 (d)** + **cyano (a)** intermolecular H-bond can serve in the stabilisation of ROY polymorphs.

2.5.2 Supramolecular Environment: Conclusions

In this section Hydrogen Bond Propensity (HBP) modelling[90] has been used to explore the supramolecular environments of the 12 unique ROY polymorphs within the CSD[25] via analysis of the predicted and observed H-bonding networks of those 12 ROY forms. HBP was able to rapidly produce a model ranking the relative favourability of the predicted H-bonding networks, expressed as a scatter plot of the mean coordination scores for each donor/acceptor atoms within the structure against the mean propensity scores of the observed H-bonds observed within that structure. This model was able to accurately identify the H-bonding network in the **YN** (QAXMEH04) form of ROY to be less than ideal and hence could contribute to metastability, a statement which is in agreement with the available thermal data for the ROY polymorphs.

Through use of the data from HBP models for the ROY structures in combination with the available thermal data, it was possible to establish a connection between relative thermal stability of a given polymorph, and the presence of an intermolecular H-bond between the **sec_amine_1 (d)** and **cyano (a)** groups within the structure. The **Y** (QAXMEH01), **O** (QAXMEH03) and **YT04** (QAXMEH12) forms of ROY all display **sec_amine_1 (d)** + **cyano (a)** H-bonds, and are all amongst the most stable ROY forms discovered to date[69,70].

2.6 2022 Crystal Structure Prediction for ROY

A 2022 study by Beran et al. produce the first crystal structure prediction (CSP) landscape for ROY which is consistent with available experimental evidence. This study utilised the 2012 CSP study of ROY performed by Vasileiadis et al[91]., which was able to predict the ten Z' = 1 polymorphs of ROY, but augmented the calculations with conformational energy corrections to the DFT lattice energies, in order to overcome the delocalisation error in the approximate density functionals. The CSP landscape produced by Beran et al. generated relative lattice energies of known polymorphs which agree well with the available experimental enthalpy data and other qualitative observations regarding polymorphic stabilities[85].

The findings from this study suggest that Z' = 1 polymorphs and, with a lesser degree of confidence, the Z' = 2 polymorphs with the most stable lattice energies have already been discovered. This study suggests that any further polymorphs to be discovered at ambient pressure would require experimental ingenuity to trap metastable forms, such as the case of **Y04** (QAXMEH53), **R18** (QAXMEH57) and

Y19 (QAXMEH60), but there may be potential to find alternative polymorphic forms at elevated pressure.

SFI analysis performed in this chapter has revealed that the **Y04** (QAXMEH53), **R18** (QAXMEH57) and **Y19** (QAXMEH60) forms display features which could hint at their metastability, particularly with respect to the intermolecular environments of these three polymorphs. FIMs analysis revealed that each of these forms contain a number of unfavourable intermolecular contacts which may serve as rationale for labelling these forms as metastable, particular when compared to the intermolecular environments presented by the **ON** (QAXMEH), **Y** (QAXMEH01) and **O** (QAXMEH03) forms.

However, it is only with the context of the intermolecular environments of both stable and metastable ROY forms, along with experimental stability information, which allow for this evaluation using SFI analysis. ROY is an ideal test case for the use of SFI to evaluate relative polymorphic stabilities, due to the relative wealth of crystallographic and stability information available. Using SFI in a discovery setting, where these data would likely be less available, would lead to less conclusive results.

Analysis of intermolecular environments using FIMs and short contacts has the potential to identify metastable forms, as is the case with the **Y04** (QAXMEH53), **R18** (QAXMEH57) and **Y19** (QAXMEH60) forms. However, the analysis performed in this chapter required the context of thermal stability data for the **ON** (QAXMEH), **Y** (QAXMEH01) and **O** (QAXMEH03) forms to evaluate, due to the qualitative nature of FIMs analysis in its current state. A means of providing a quantitative measure of the favourability of intermolecular short contacts based on their proximity to FIMs hotspots could provide a more robust evaluation of intermolecular environment using just SFI tools. However, in its current state, SFI analysis requires the context of experimental stability and crystallographic data to draw meaningful conclusions.

2.7 Conclusions

In this section, SFI analysis was used alongside available thermal data in order to evaluate the relative stabilities of the 12 ROY polymorphs in the CSD. Analysis of the intermolecular environment of ROY polymorphs using MGC highlighted the metastable **Y04** (QAXMEH53) form as displaying an unusual θ_{thio} torsion. When viewed in context of the intermolecular gas-PES for ROY calculated by Cruz-Cabeza et al. in 2014[86] the unusual θ_{thio} torsion displayed by **Y04** corresponds to a position close to the maximum of the gas-PES curve for ROY. This additional context could be used to label the **Y04** form as potentially metastable.

Analysis of the intermolecular environments of selected stable and metastable ROY forms using a combination of FIMs and intermolecular short-contacts revealed noticeable differences in the intermolecular environments of stable ROY forms to those of metastable ROY polymorphs. Whilst analysis of the supramolecular environments of the ROY polymorphs through the use of HBP modelling revealed that the stable Y , O and YT04 forms of ROY all display an intermolecular **sec_amine_1 (d)** + **cyano (a)** H-bond.

With the large amount of crystallographic data available for the ROY polymorphs, and the available experimental stability data, it is possible to make an assessment of a given polymorph as o whether it is likely to be stable or metastable. However, it is this additional contextual data which provide the basis for evaluation. Without additional context, the intermolecular environment data provided by FIMs and short contacts could be used to provide a qualitative assessment of relative polymorphic stability, but a confident evaluation would require a more robust, quantitative evaluation of the intermolecular environment of a given polymorph. In its current state, SFI analysis alone is unlikely to be able to de-risk polymorphic forms, and significant additional contextual crystallographic and experimental stability data are necessary in order to make confident evaluations of relative polymorphic stability.

Chapter 3 Design, Synthesis, SFI Analysis and Polymorph Screening of a Set of Close Structural Derivatives of ROY

3.1 Aims and Objectives

In Chapter 2 SFI analysis was used to identify a number of molecular features of the ROY scaffold which may play a role in the high number of polymorphic forms accessible to the ROY scaffold. In this chapter a set of 12 functional derivatives of the ROY scaffold were designed to probe the influence of the aforementioned molecular features on the polymorphism of the 12 ROY derivatives.

After their design, synthesis and characterisation, crystals of the 12 ROY derivatives were analysed using SFI. This focus of this SFI analysis was to map the changes in the intra-, inter-, and supramolecular environments of the ROY derivatives relative to the ROY scaffold analysed in Chapter 2.

Finally, a low-/medium-throughput polymorph screen was designed and applied to the 12 ROY derivatives synthesised in this chapter. The goal of these screens is to provide evidence of the ability of the ROY scaffold to produce functional derivatives which display polymorphism, a key factor in confirming the status of the ROY scaffold as a polymorphophore.

3.2 Introduction

The extreme degree of polymorphism it displays, along with its status as a polymorphophore, make the ROY scaffold an ideal starting point to evaluate the exploratory power of the SFI workflow. Through Chapter 2 it was demonstrated that the SFI workflow could be used to explore the crystal structures of the unique ROY polymorphs within the CSD at different length scales. This allowed for the identification of several intra-, inter- and supramolecular features which could be used to help rationalise the relative polymorphic stabilities of the ROY polymorphs.

In an attempt to establish a Structure-Polymorphism Relationship (SPR) for the ROY polymorphophore a set of 12 compounds were designed, synthesised, characterised and subjected to a small-scale polymorph screen. The design of these 12 compounds was guided by the SFI analysis performed in Chapter 2, with structural modifications designed to influence the same inter-, inter- and supramolecular factors identified during the SFI analysis of the ROY polymorphs.

The goal of this section is to document the design and synthesis of these 12 structural derivatives of ROY, along with an SFI exploration of their crystal environments in order to document the observed changes driven by those

structural modifications. Finally, the SFI exploration of the 12 ROY derivatives is compared to the results of the small-scale polymorph screen in order to explore the SPR of the ROY polymorphophore

3.3 Design and Synthesis of ROY Derivatives

The structural modifications made to the ROY scaffold in this Chapter were designed to affect the most influential intra-, inter- and supramolecular environmental factors identified in Chapter 2. The factors include interring torsion (herein referred to as θ_{thio}), which was identified using Mogul Geometry Check[80] to correlate with polymorphic stability. Moreover, it was found that structures which were Matched-Molecular Pairs[92] (MMPs) of ROY behave more similarly to ROY than structures with larger structural/molecular differences, thus derivatives designed in this section were designed to be MMPs of the ROY scaffold.

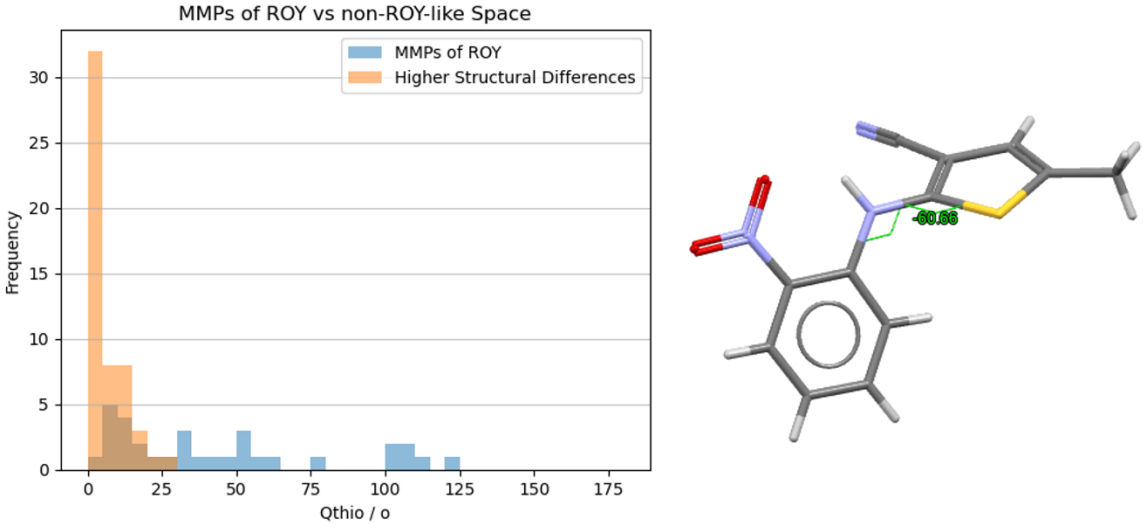


Figure 63; (Left) Distribution of θ_{thio} values of unique structures containing similar structural motifs to ROY. θ_{thio} values of MMPs of ROY are highlighted in blue, values of compounds with higher structural differences are highlighted in orange. (Right) the θ_{thio} torsion highlighted in green for the Y19 (QAXMEH60) form of ROY

Evaluation of the intermolecular environments of stable vs unstable ROY polymorphs using Full Interaction Maps[93] (FIMs) identified that the nitrile and nitro groups were the most influential on polymorphic stability, along with the bridging amine and aromatic proton groups most commonly found as interaction partners with these groups. As such, these were key targets for modification in the set of ROY derivatives.

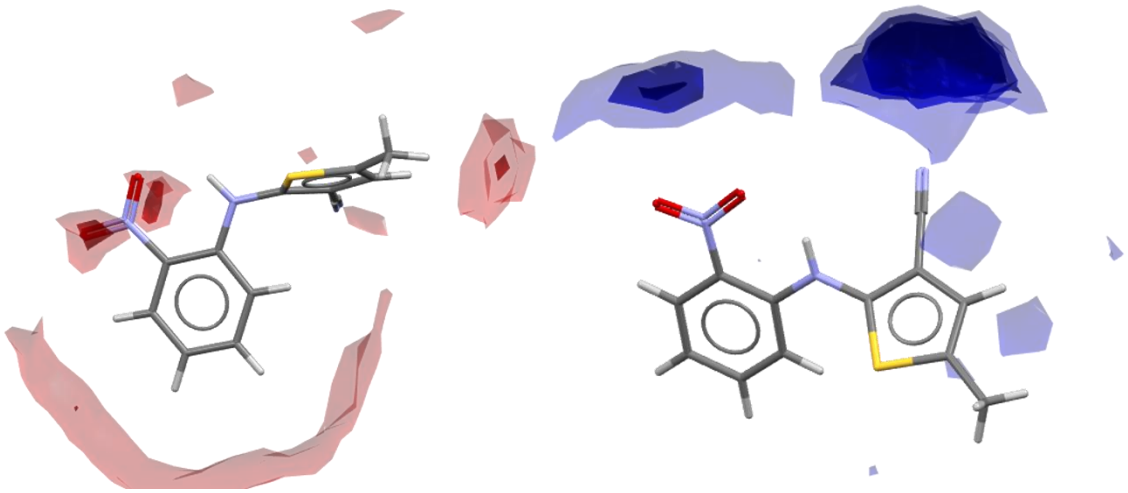


Figure 64; (Left) Carbonyl oxygen (red) FIMs hotspot calculated for the Y (QAXMEH01) form of ROY. (Right) uncharged NH hydrogen (blue) FIMs hotspot calculated for the ON (QAXMEH) form of ROY.

Finally, HBP analysis of the 12 unique ROY polymorphs within the CSD found that the presence of intermolecular H-bonds between bridging amine proton and the nitrile nitrogen of the ROY scaffold were indicative of stable forms, whereas H-bonds involving other acceptor groups (such as the nitro oxygen) are indicative of more metastable forms. This made modification of the number and identities the of H-bonding partners present within the ROY scaffold a key factor to modify in order to influence the supramolecular environments of the ROY derivatives.

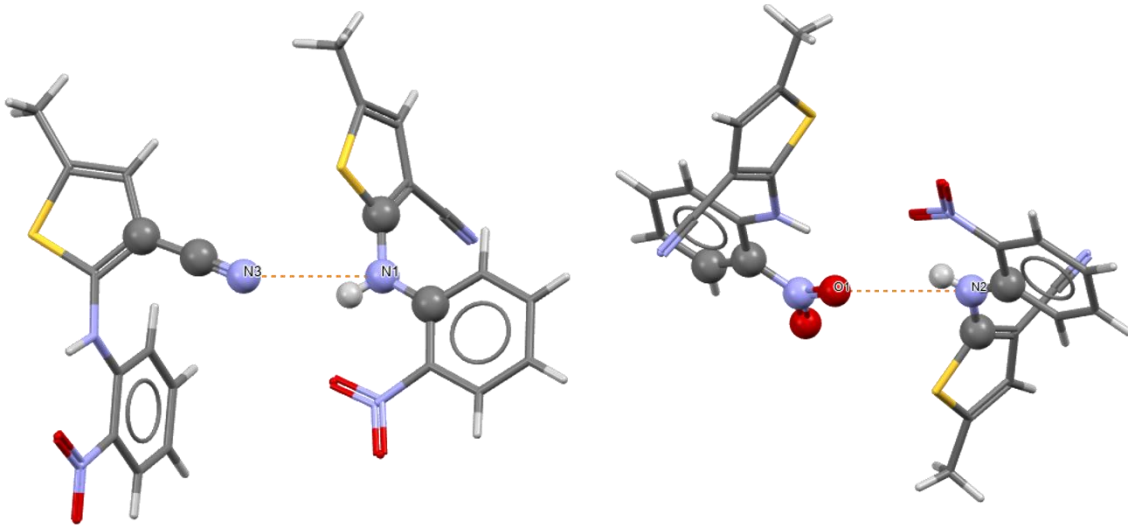


Figure 65; (Left) intermolecular H-bond observed between the cyano and sec_amine_1 groups from the Y (QAXMEH01) form of ROY. (Right) intermolecular H-bond observed between the ar_nitro and sec_amine_1 groups from the YN (QAXMEH12) form of ROY.

3.3.1 Chloro-/Methyl-Exchanged Derivatives

Alongside the SFI analysis of the ROY scaffold from Chapter 2, examples of studies probing the polymorphism of other polymorphophores from the literature

were examined. In particular, a 2015 study into the SPR of the fenamate polymorphophore by Lopez-Meijas proved useful[94]. In this study, Lopez-Meijas and coworkers substituted aromatic C-H bonds of benzene rings present in the fenamate structure with C-Cl, C-Br and C-CH_3 bonds in order to create a library of MMPs of the fenamate polymorphophore. This study looked to probe the 2 and 3 positions of the fenamate benzene ring for their impact on the resultant polymorphism of the substituted structure.

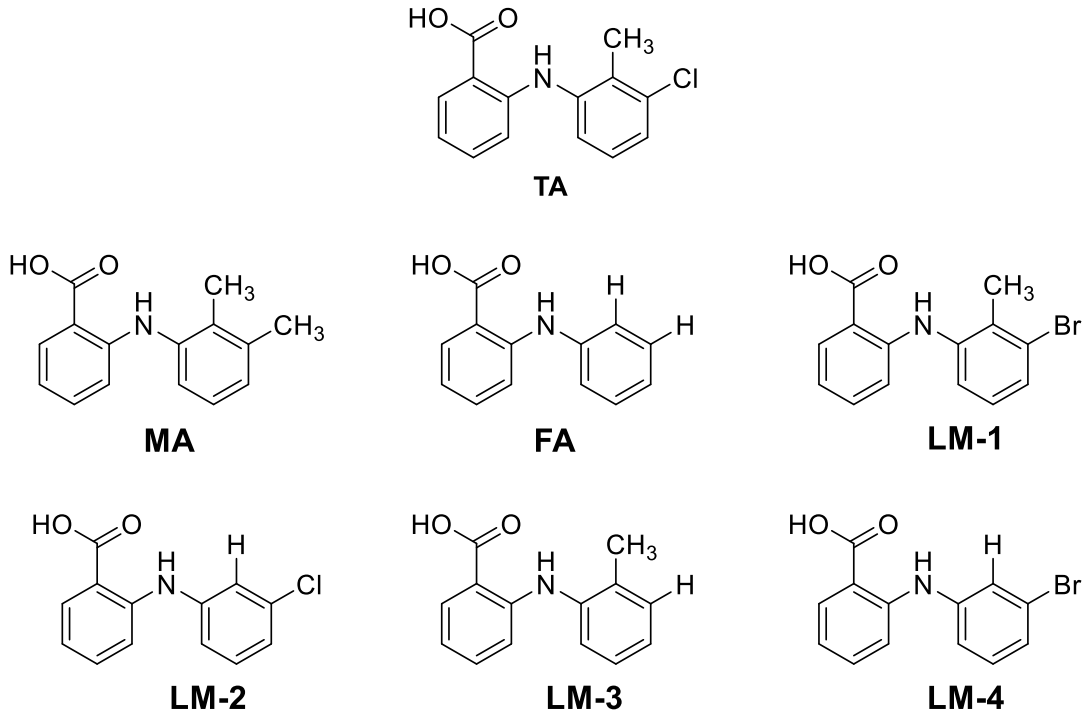


Figure 66; Structures for the seven structures investigated by Lopes-Mejias et al in their 2015 study into the fenamate polymorphophore. Tolfenamic acid (TA), mefenamic acid (MA) and fenamic acid (FA) were structures previously reported in the literature, whereas structures LM-1 – LM-4 were synthesised for the 2015 study.

There is insufficient evidence to identify any particular aromatic protons in the ROY structure that are significant enough to warrant as deep a study as the Lopes-Meijas study provides. However, the chloro/methyl exchange from the 2015 study provides a useful means to probe the influence of each of the aromatic protons in the ROY structure. The chloro-/methyl-exchange allows for the position of the aromatic proton to be probed, alongside the nature of the contacts likely to form. The relatively electron-deficient aromatic proton can be compared to the relatively electron-rich methyl proton, or the similarly electron-rich lone pair of the chlorine atom. Figure 67 contains the structures of the chloro/methyl ROY derivatives synthesised in this section.

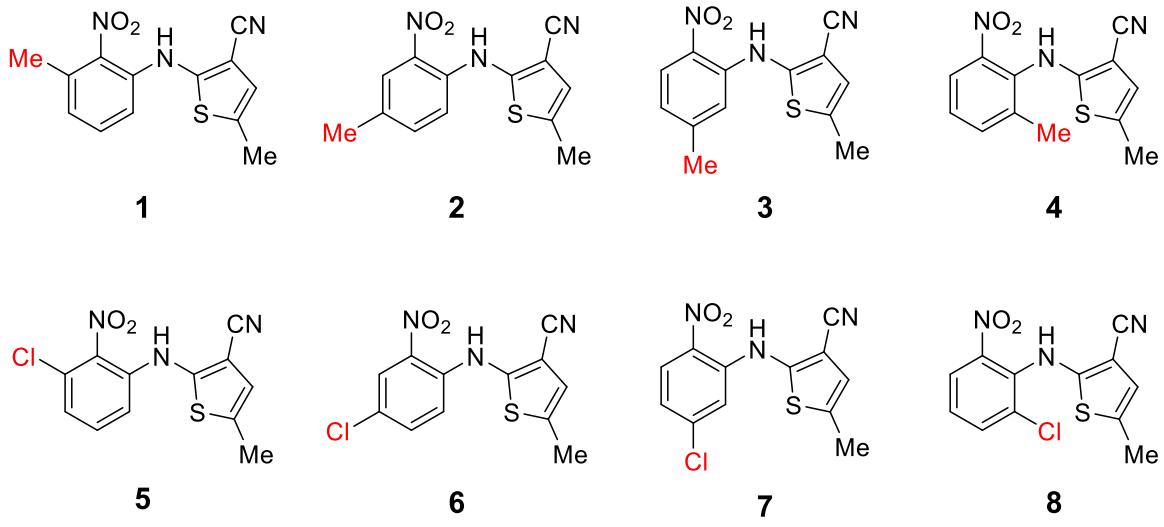


Figure 67; structures for the eight chloro/methyl exchanged ROY derivatives synthesised during this section, along with their corresponding reference number used to refer to each structure. Modifications made to the ROY structure are highlighted in red.

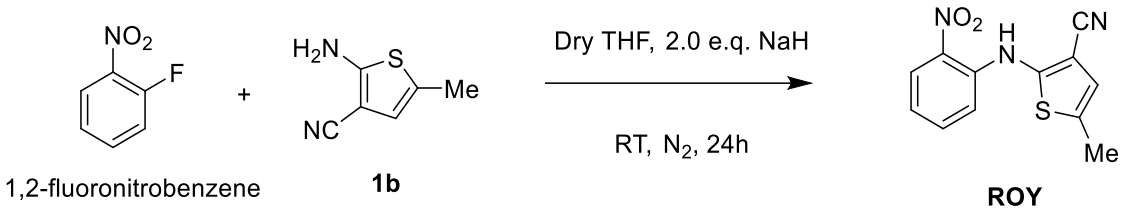
3.3.1.1 Method Development

The nature and levels of different impurities present within a solution can have a significant impact on the result of any crystallisation experiment. As one of the goals of this work is to perform a polymorph screen in order to explore and compare the polymorphic profile of the twelve ROY derivatives experimentally, it is important for us to attempt to minimise the differences in the impurity profiles between the twelve compounds. One way this was achieved was to minimise the number of different synthetic routes used to synthesise the twelve structures. It was found that ten of the twelve structures (83%) could be synthesised using a modified version of the route which has been used to produce the ROY structure previously within the literature. This reaction consists of a nucleophilic aromatic substitution (S_NAr) reaction between 1,2-fluoronitrobenzene and 5-methyl-2-aminothiophenecarbonitrile in the presence of a base (NaH) under dry conditions[68].

In order to assess the reproducibility of this route, the Shevchenko route was first repeated under identical conditions to produce the unsubstituted ROY structure. This route is provided in Figure 66. Pure product was isolated via recrystallisation from IPA.

The reported yield for this reaction was 25%, whilst the average yield from the repeat experiments performed for this work was found to be around 18%. The low yield for this route was highlighted as a potential issue, suggesting further work was necessary to optimise the reaction if it were to be used to produce ROY derivatives at scale. Nonetheless, this route was seen as a simple, quick and

reproducible route to produce the ROY structure, which could potentially be used as a base for the production of our structural ROY derivatives.

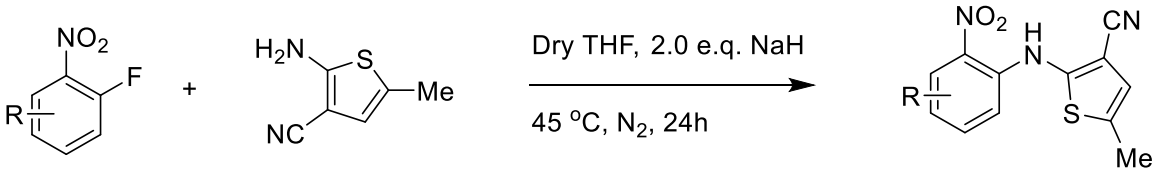


1,2-fluoronitrobenzene **1b** **ROY**

Yield (Schenchenko et al, 2005) = 25%
Yield (This work) = 18%

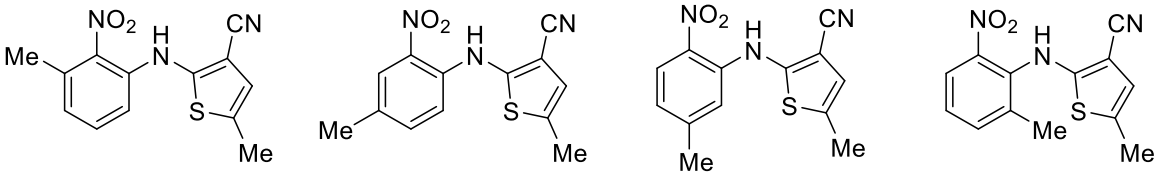
Figure 68; Synthetic route taken from Schevchenko et al. (2005) describing an SNAr-based route to the one step synthesis of ROY from 1,2-fluoronitrobenzene and 5-methyl-2-aminothiophene carbonitrile. The reported yield from Schevchenko et al. (performed at 3 mol scale) along with the yield reported from this work (performed at 1 mmol scale) are also given.

Satisfied with the ability of this route to produce the ROY scaffold, the synthesis was modified in order to produce our chloro- and methyl-ROY derivatives. The respective chlorinated/methylated fluoronitrobenzene substrate was substituted for 1,2-fluoronitrobenzene, yielding in ROY derivatives **1** – **8** in yields between 11% - 20%.



1a - 8a 1b **1 - 8**

R = CH_3 Cl

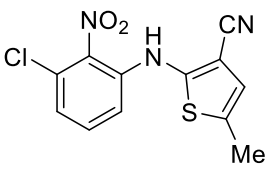


1
21.8 mg, 10 %

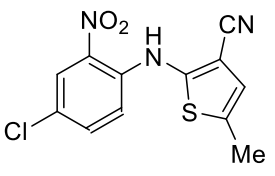
2
18.7 mg, 7 %

3
125.2 mg, 23 %

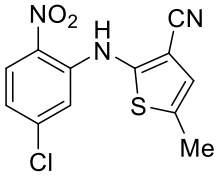
4
33.1 mg, 12 %



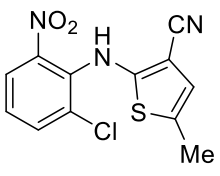
5
35.1 mg, 12 %



6
12.5 mg, 5 %



7
19.4 mg, 7 %



7
29.1 mg, 9 %

Figure 69; (Top) General reaction scheme for the SNAr synthesis of the eight chloro/methyl ROY derivatives, adapted from Schevchenko et al. (2005), where R is a chloro or methyl group, positioned at one of the four unsubstituted positions around the benzene ring. (Bottom) Chemical structures, yields and pseudonyms for the eight chloro/methyl ROY derivatives synthesised via the Schevchenko route.

Despite success in generating the desired structures, the low yields achieved using the Schevchenko S_NAr route meant an alternative route to generating the functionalised nitrobenzene moieties was necessary. Fortunately, due to it being a key intermediate in the synthesis of olanzapine, there are a number of available routes. One route, taken from a Leya-Perez et al. study into the industrial production of olanzapine, utilises N,N-dimethylformamide (DMF) as a solvent along with a weaker inorganic base in KOH to produce ROY in good yields[95]. It was this route which was modified in order to produce ROY derivatives **1** – **8** on the 1000 mg scale required for the subsequent polymorph screen. Figure 70 contains the updated general route utilised to generate gram-scale levels of ROY derivatives **1** – **8**.

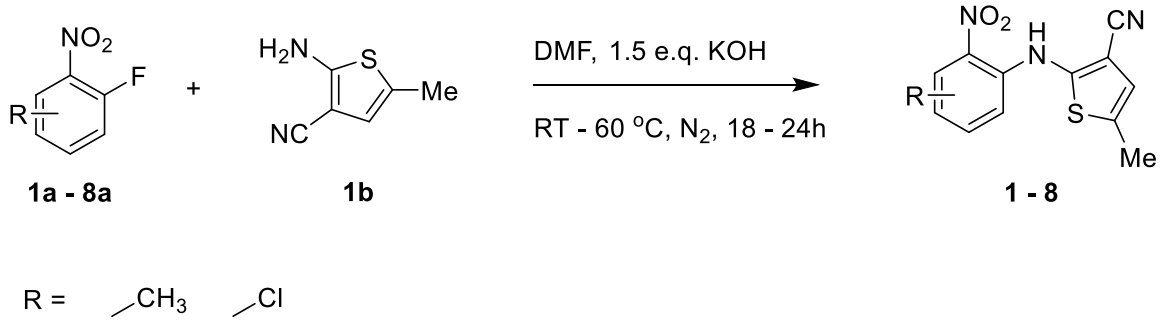
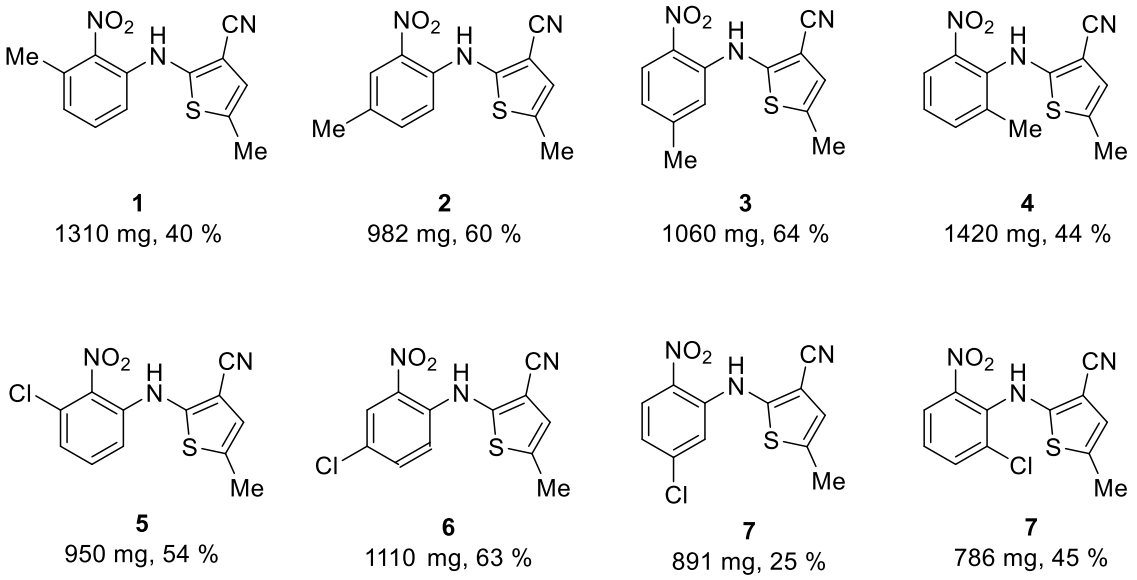


Figure 70; Updated general scheme for the production of ROY derivatives 1 – 8, utilising route adapted from Leyva-Perez (2010)[95].

Alongside the alterations to the route used, small adaptations were made to the setup and purification of the reaction to aim for higher yields and purity of final product. The first change was in the addition of the reagents **1a** – **8a** and **1b**, which were combined separately to the KOH solution and subsequently added dropwise to the KOH solution. This resulted in a characteristic blue hue of the reaction mixture after around five minutes post addition, which was found to be indicative of successful addition. Secondly, to ensure maximum removal of the DMF solvent, an additional washing step was introduced during the work up. Eight washes of 1M LiCl (aq) solution, in total equating to five times the initial volume of DMF used, were employed to remove the bulk of DMF from the crude product before a flash column using 1:4 EtOAc/hexane column solvent yielded pure product. Overall, this resulted in improved yields of ROY derivatives **1** – **8**, generating sufficient quantity and purity of each derivative for the subsequent polymorph screen.



1
1310 mg, 40 %

2
982 mg, 60 %

3
1060 mg, 64 %

4
1420 mg, 44 %

5
950 mg, 54 %

6
1110 mg, 63 %

7
891 mg, 25 %

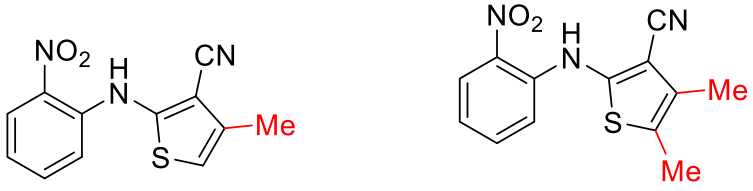
7
786 mg, 45 %

Figure 71; structures, yields and quantities produced of chloro/methyl ROY derivatives 1 – 8. All derivatives were produced using the method adapted from Leyva-Perez[95].

3.3.2 Aminothiophene Derivatives

Originally, the aminothiophene ROY derivatives were designed as part of the chloro-/methyl-exchanged ROY derivatives looking to probe the role of aromatic protons in the stabilisation of MMPs of ROY. The lack of economical options to obtain the aminothiophene substrates necessary for the S_NAr synthesis of the chloro-/methyl-ROY derivatives, along with the significant roadblocks encountered during the synthesis of those aminothiophene substrates, necessitated that the aminothiophene chloro-/methyl-ROY derivatives are discussed separately.

The aminothiophene ROY derivatives were designed to probe the effects of substituents at the 4 and 5 position of the aminothiophene ring of the ROY polymorphophore. There were originally more aminothiophene ROY derivatives planned, but time constraints and synthetic issues meant that only two were developed. Derivative **9** features a ROY scaffold with a methyl group at position 4 and a proton at position 5, whilst derivative **10** features methyl groups at both positions 4 and 5 of the thiophene ring.



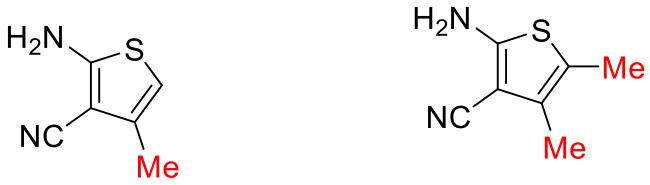
9

10

Figure 72; structures for the two aminothiophene ROY derivatives designed and synthesised for this section. Changes to the ROY scaffold are highlighted in red.

3.3.2.1 Method Development

The primary difficulty posed by the aminothiophene derivatives is the production of the 4-substituted and 4,5-substituted aminothiophene substrates required for the S_NAr step used to produce their respective ROY derivative. The two substrates in question (labelled **9b** and **10b**, respectively) can be found in Figure 73.



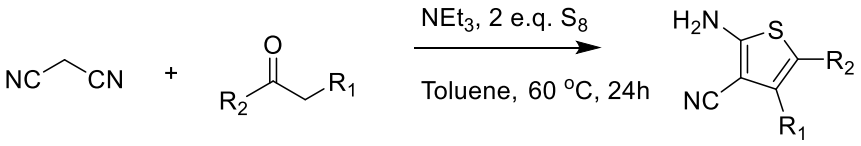
9b **10b**

Figure 73; structures for the two substituted aminothiophene substrates designed and synthesised for use in the synthesis of ROY derivatives 9 and 10.

As with previous analogues, the literature was searched for previous syntheses of ROY and/or olanzapine in order to establish the routes typically employed to access the 5-methylaminothiophene substrate required in their synthesis. The most commonly found route to produce the 5-methyl-2-amino-3-thiophene carbonitrile substrate used in the synthesis of ROY/olanzapine is based on the second generation of the Gewald aminothiophene synthesis, discovered Gewald et al. in 1966[96]. This route utilises malononitrile, the corresponding carbonyl reagent and ocstasulfur along with a base in order to generate the desired aminothiophene species and is summarised in Figure 74.

Gewald Aminothiophene Synthesis

Gewald, K. et al. (1966)



Proposed Mechansim:

Puetrova, Z. et al. (2010)

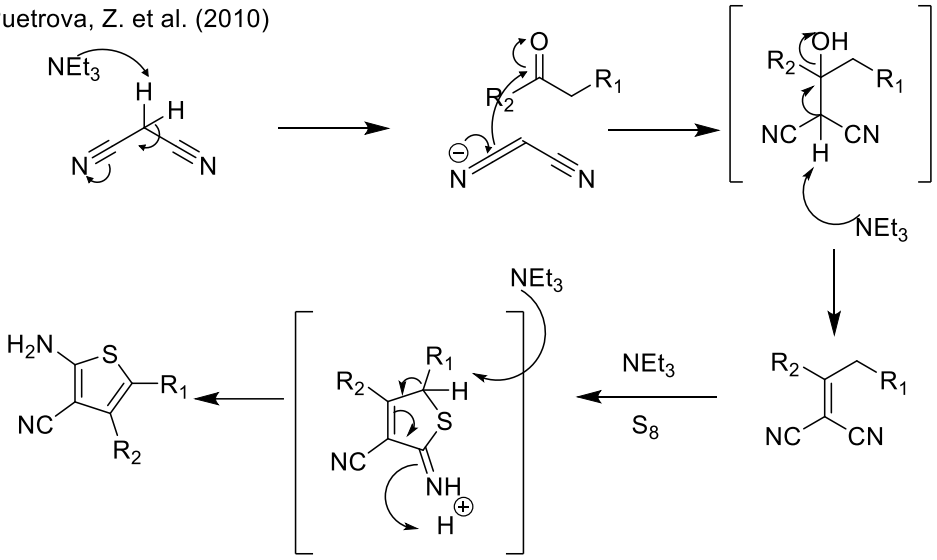


Figure 74; (Top) General reaction scheme for the 2nd generation Gewald aminothiophene synthesis.[96] (Bottom) Proposed mechanism for the 2nd generation Gewald synthesis (adapted from [97]).

The reaction proceeds via a Knoevenagel condensation between malononitrile and the carbonyl species, mediated through the base species added to the reaction mixture. This allows for the generation of aminothiophene substrates which can be modulated at the 4 and 5 position through the substituents present in the carbonyl species. It was theorised that this single route could be utilised with commercially available chlorinated and methylated carbonyl species in order to access the aminothiophene substrates necessary for the production of **9b** and **10b**.

A number of reactions were performed in order to ascertain the conditions necessary to produce the necessary amounts of **9b** and **10b** via the route outlined in Figure 74 utilising a range of different solvents and temperatures. A solvent of dry tetrahydrofuran (THF) was identified as the most promising solvent at this stage, despite its low overall yields, and thus this reaction was scaled-up to 30 mmol in an attempt to generate sufficient levels of **9b** and **10b**.

There was a degree of success to this synthesis, producing enough **9b** and **10b** to allow for synthesis and characterisation of the ROY derivatives **9** and **10** for the first time. However, the overall efficiency of the reaction was poor and

deemed insufficient for the production of gram-scale amounts of ROY derivatives **9** and **10**, and an alternative route was required.

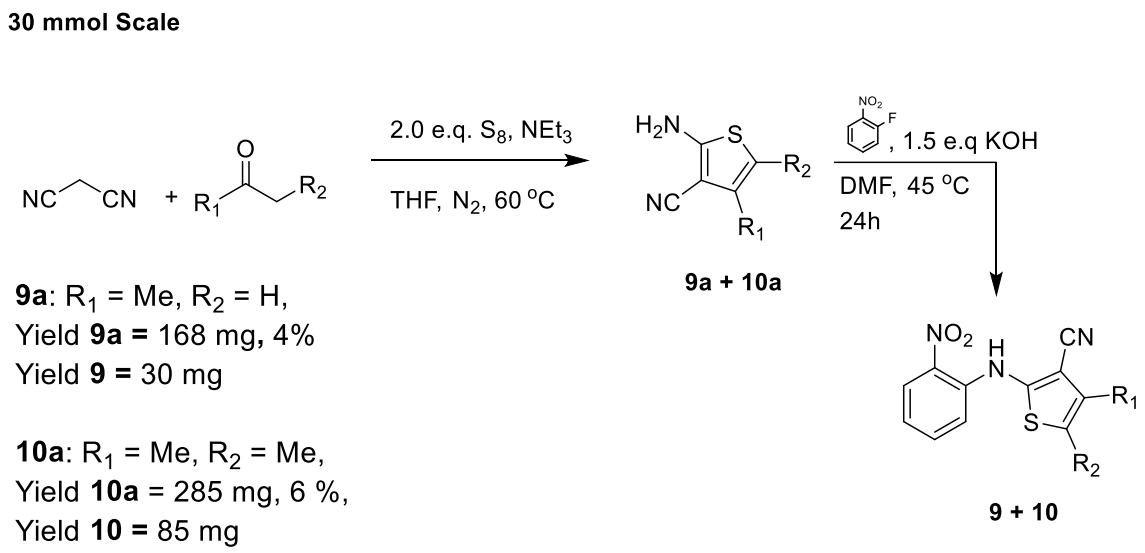


Figure 75; summary of the 30 mmol Gewald aminothiophene syntheses, and subsequent SNAr syntheses of aminothiophene ROY derivatives 9 and 10.

One route that showed promise was proposed by Tao Wang et al. in 2010, which utilised a catalytic amount of L-proline in order to produce substituted aminothiophene species in good yields[98]. This route utilised L-proline as an organocatalyst in place of a basic species, alongside DMF solvent and mild heat, and was able to produce 2-aminothiophene species with a diverse set of modifications to the 4- and 5-positions in high yields. A summary of the L-proline catalysed syntheses from Wang et al. is found in Figure 76.

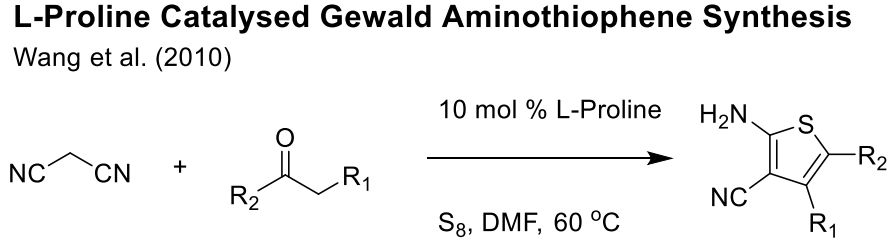


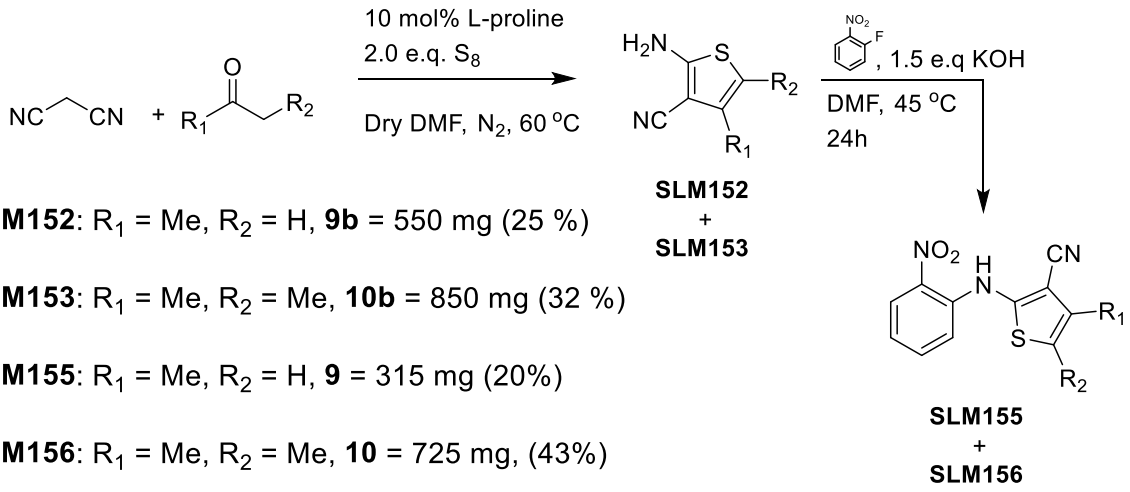
Figure 76; L-proline catalysed Gewald aminothiophene reported by Wang (2010)[98], along with selected substrates and yields.

The simplicity and efficiency of this route made it highly appealing, and two small-scale syntheses of **9b** and **10b** were performed utilising these conditions. These experiments were found to be a major improvement over previous aminothiophene syntheses, with yields improving significantly over the NEt_3-mediated route discussed previously. Thus, the L-proline catalysed route was selected as the primary route for the production of functionalised aminothiophene species going forward.

In addition to the improved efficacy achieved utilising the L-Proline catalysed route, a small adjustment was made to the extraction and purification step at the end of the reaction. The multi-step purification process used previously was replaced with a wash with 1M LiCl solution was used to remove the DMF, followed by flash column chromatography using 100:1:0.01 DCM/MeOH/NEt_3 as the column solvent. This achieved the same level of purity achieved using the original route using much fewer steps.

3.3.2.2 Synthesis of Aminothiophene ROY Derivatives

Synthesis of aminothiophene ROY derivatives **9** and **10** was achieved by first synthesising aminothiophene substrates **9b** and **10b**, using the L-proline catalysed route reported by Wang et al. These syntheses were performed at 12 mmol scale and produced 550 mg **9b** and 875 mg **10b** respectively. These aminothiophene substrates were combined with 1,2-fluoronitrobenzene under basic conditions in order to produce their respective aminothiophene ROY derivatives. Figure 77 summarises the final synthetic route used to generate aminothiophene ROY derivatives **9** and **10**.



SLM152: R_1 = Me, R_2 = H, **9b** = 550 mg (25 %)

SLM153: R_1 = Me, R_2 = Me, **10b** = 850 mg (32 %)

SLM155: R_1 = Me, R_2 = H, **9** = 315 mg (20%)

SLM156: R_1 = Me, R_2 = Me, **10** = 725 mg, (43%)

Figure 77; General reaction scheme and conditions used for the production of aminothiophene substrates 9b and 10b and subsequent synthesis of aminothiophene ROY derivatives 9 and 10. Step 1 performed at 12 mmol scale and adapted from[98]. Step 2 performed at 4 – 6 mmol scale and adapted from[95].

3.3.3 Amino-ROY

Amino-ROY was designed to be a MMP of the ROY scaffold in which the nitro-group attached to the benzene ring of ROY is selectively reduced to the primary amine. Amino-ROY (herein referred to as **11**) is unique amongst the ROY derivatives synthesised in this section as it is one of only two structures which have been previously reported in the literature[95], along with having an entry in the

CSD (Refcode = XABXEF) Structures and CSD refcodes for ROY and **11** can be found in Figure 78.

ROY
QAXMEH

11
XABXEF

Figure 78; structures of ROY and amino ROY derivative 11, with structural differences highlighted

The nitro-group in ROY plays a key role in restricting the intramolecular geometries available to the structure via the formation of a strong, six-membered ring intramolecular H-bond with the bridging amine proton. As this intramolecular motif is present in all known ROY polymorphs it functionally locks the orientation of the nitrobenzene moiety in all ROY structures. The result of the presence of this six-membered ring intramolecular H-bond is that the orientation of the nitrobenzene moiety remains constant across all ROY polymorphs within the CSD, whilst the thiophene ring is more labile and free to adopt alternate orientations.

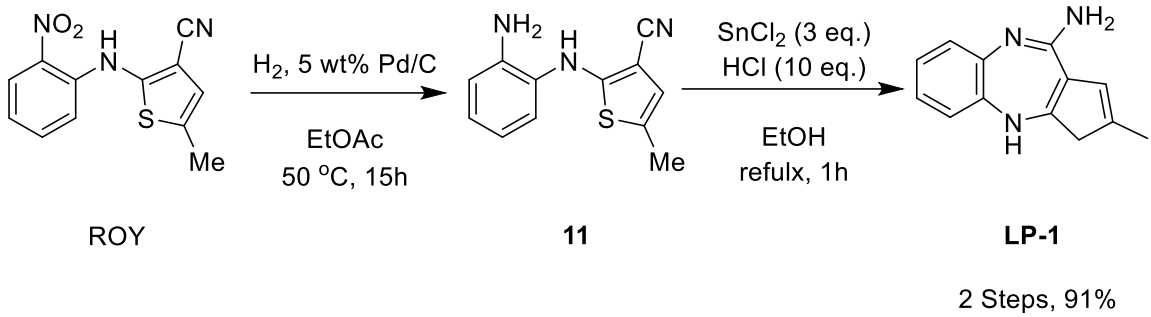
As was discussed in Chapter 2.1, this variability in the relative position of the thiophene ring plays a role in the nature, number and quality of the intermolecular interactions which can be formed by the ROY scaffold, and hence is theorised to play a role in facilitating the large number of polymorphic forms accessible to the ROY scaffold.

3.3.3.1 Method Development

As **11** is a structure which has been previously reported in the literature, the selection process for deciding upon which route to use is significantly easier. The selected route used is this section comes from the 2010 study into the industrial production of olanzapine from Leyva-Perez et al., which was also used to inform the routes used in sections 3.1.1 and 3.1.2[95]. Through this route Leyva-Perez et al. utilised a catalytic hydrogenation reaction, using a 10 mol% powder of palladium supported on activated charcoal (Pd/C) and hydrogen gas to reduce the nitro group present in ROY to produce **11** as part of a two-step process to produce olanzapine intermediate **LP-1**, with an overall yield of 91%. Reaction schemes, structures and yields for the production of **LP-1** can be found in Figure 79.

Leyva-Perez et al. (2010)

Taken from reported Olanzapine synthesis



ROY **11** **LP-1**

2 Steps, 91%

Figure 79; (Top) a section of the reported olanzapine synthesis taken from a 2010 study from Leyva-Perez et al., including the reported yield from the two steps involving the Pd/C hydrogenation of the ROY scaffold, taken from[95].

The high yield and simplicity provided by the Pd/C reduction made this route highly appealing. Small scale hydrogenations were performed to test the efficacy of the Pd/C route, which were successful, leading to generation and characterisation of **11**. Two repeats of this experiment were subsequently performed, both at 3.5 mmol scale, in order to produce sufficient stock of **11** for the polymorph screen. A summary of the Pd/C reduction of ROY to **11** discussed in this section, including yields, can be found in Figure 80.

This work

Lab Reference	Yield
SLM144	83%

ROY **11**

Figure 80; Reaction scheme for the 3.5 mmol scale Pd/C reduction of ROY to 11, taken from Leyva-Perez [95]. Yields are highlighted

3.3.4 ROY Aldehyde

The final ROY derivative designed and synthesised for this section (herein referred to as **12**) is a structure in which the nitrile group in ROY was converted to the corresponding aldehyde via selective partial reduction of the nitrile group, followed by work up with H_2O. Figure 81 contains structures of ROY and **12**, highlighting structural differences.

As was the case with derivative **11**, the structural modifications resulting in derivative **12** were designed to impact the intra-, inter- and supramolecular environments of the ROY scaffold simultaneously, and as such, the introduction

of an aldehyde group in place of the nitrile group at the 3 position in the ROY scaffold was highly appealing. The position of the oxygen atom in the aldehyde group is such that an additional six-membered ring intramolecular H-bond between the aldehyde and the bridging amine is likely to form. Coupled with the six-membered ring H-bond observed between the nitro group and the bridging amine, this modification is likely to further restrict the accessible θ_{thio} values. Info regarding θ_{thio} in the observed form(s) of **12** should shed light on the influence θ_{thio} has on the underlying polymorphism of the ROY polymorphophore.

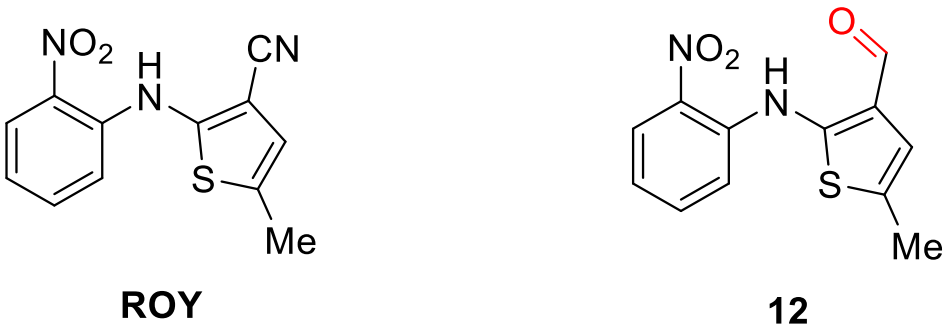


Figure 81; structures of ROY and ROY aldehyde derivative 12, with structural differences highlighted.

Another factor in the selection of the aldehyde group in particular is the potential impact the group may have on the intermolecular environment of the ROY polymorphophore, as highlighted by FIMs hotspots. Chapter 2.2 demonstrated the importance of the nitrile group on the polymorphic stability of ROY polymorphs, with stable forms such as **ON** (QAXMEH) displaying favourable nitrile environments, whereas metastable forms such as **R18** (QAXMEH57) display unfavourable intermolecular H-bonds. Replacing the nitrile group with an aldehyde group introduces an oxygen atom with altered electron density to the nitrile nitrogen atom in ROY, along with a somewhat acidic proton, which should significantly alter the FIMs hotspots and observed intermolecular environment/contacts formed by forms of **12**.

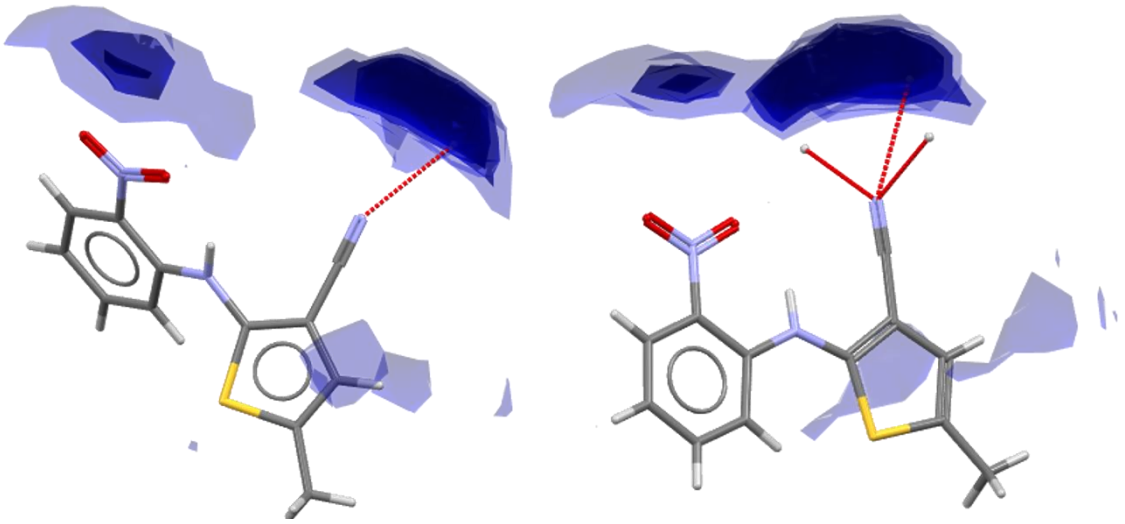


Figure 82; (Left) uncharged NH proton (blue) FIMs hotspot calculated for the ON (QAXMEH) form of ROY, along with the observed intermolecular contact. (Right) uncharged NH proton (blue) FIMs hotspot calculated for molecules B of the R18 (QAXMEH) form of ROY, along with the observed intermolecular contacts.

Finally, Chapter 2.3 demonstrated that the supramolecular environment observed for the 12 unique ROY polymorphs is primarily dictated by the presence (or absence) of intermolecular H-bonds, with H-bonds observed between the bridging amine and the nitrile group being indicative of stable forms, whereas intermolecular H-bonds formed with other acceptor groups being indicative of metastable forms. Replacing the nitrile with an aldehyde in **12** introduces another H-bond donor, whilst maintaining the same number of H-bond acceptors overall, whilst the intramolecular H-bond predicted to form would decrease the likelihood of H-bonds forming between the bridging amine and the aldehyde oxygen. These modifications are used to probe the influence of intermolecular H-bonds on the stability of crystals containing the ROY polymorphophore.

3.3.4.1 Reduction of ROY to ROY-Aldehyde *via* Diisobutylaluminium Hydride/H_2O

In order to partially reduce the nitrile group present in the ROY scaffold, mild reductive conditions are required. One species capable of selectively reducing a nitrile group is the Lewis acidic diisobutylaluminium hydride (DIBAL-H). The DIBAL-H group coordinates to the nitrile nitrogen, which then allows for the migration of a hydride to the nitrile carbon, resulting in an aluminium-bound imine species. This species can then be worked up through the addition of H_2O in order to yield the corresponding aldehyde. The general mechanism for the DIBAL-H reduction of a nitrile species can be found in Figure 83.

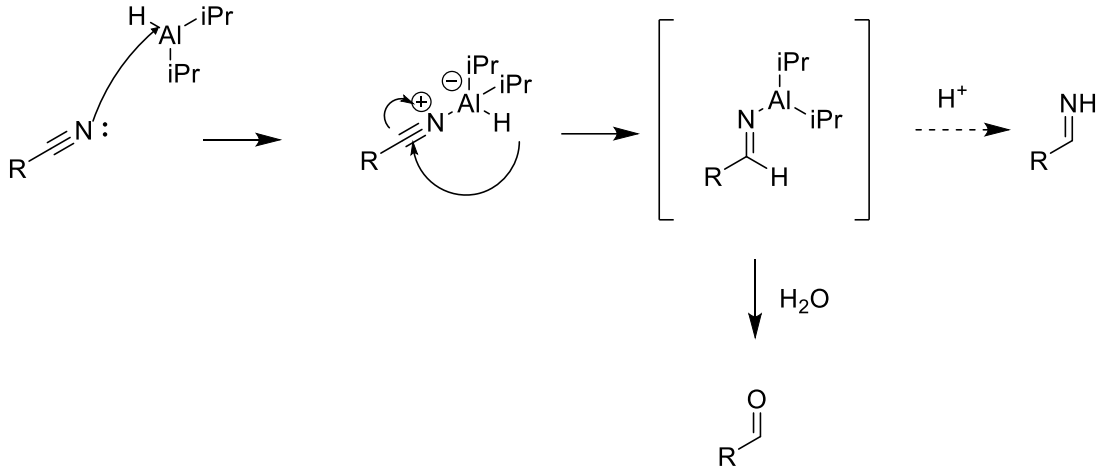
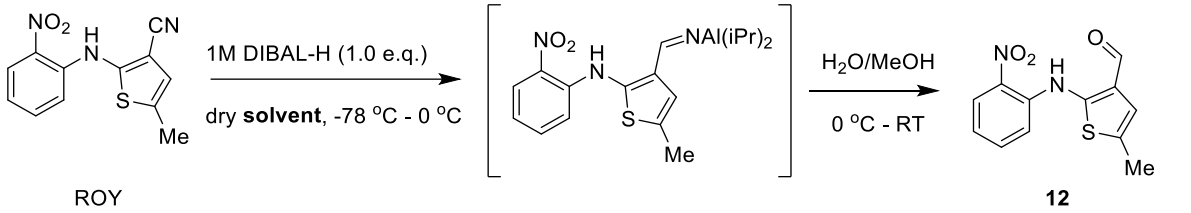


Figure 83; General mechanism for the reduction of a nitrile group via DIBAL-H.

In order to evaluate the applicability of DIBAL-H reduction in the production of **12** a number of small scale DIBAL-H reductions of the ROY scaffold were performed using a range of solvents. ROY was dissolved in the selected solvent and cooled to -78 ºC using an acetone/dry ice slurry. Once temperature stabilised DIBAL-H 1M solution in hexanes (1.0 eq.) was added dropwise to the ROY solution. Upon completion of the DIBAL-H addition, the temperature was maintained at -78 ºC for one hour before the acetone/dry ice slurry was replaced with an ice/water slurry and the reaction was allowed to slowly warm to -10 ºC. Reaction progress was monitored via TLC. Once depletion of starting material was observed the reaction was quenched using a 4:1 ratio H_2O/methanol mixture. ROY aldehyde derivative **12** was isolated as dark red needles via flash column chromatography using a 2:1 EtOAc/hexane column solvent. Figure 84 contains a summary of the small-scale DIBAL-H reduction experiments.



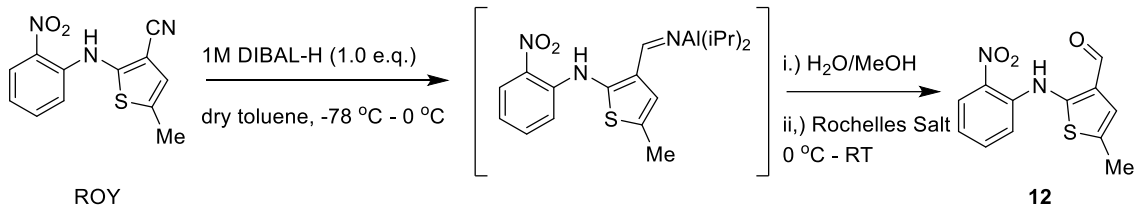
Solvent = Tetrahydrofuran (dry), Yield = 73 mg (28 %)
Solvent = Toluene (dry), Yield = 160 mg (61 %)
Solvent = Dichloromethane (dry), Yield = 66 mg (25 %)

Figure 84; summary of the 0.5 mmol DIBAL-H reductions of ROY to 12. Solvents used and yields are highlighted.

The DIBAL-H route was shown to be a reliable method for the production of **12** and a repeat synthesis was planned at a larger scale. It was noted that, despite TLC indicating full conversion of starting material, yields observed for the small-scale DIBAL-H reductions were relatively low. One possible reason for this was

a loss of material during the work-up and extraction phase. In an attempt to rectify this issue, an additional washing step was added to the work-up, with 1M sodium potassium tartrate (Rochelle's salt) solution added after the addition of H_2O/methanol solution. Rochelle's salt coordinates to any aluminium species in solution, allowing for efficient and effective removal of aluminium species via the aqueous layer.

With the revised route established, two 3.5 mmol (1000 mg) scale DIBAL-H reductions of ROY were performed. These syntheses allowed for the generation of sufficient stock of **12** for the polymorph screen. Figure 85 contains a summary of the 3.5 mmol scale DIBAL-H reductions of ROY to yield ROY aldehyde derivative **12**.



Solvent = Toluene (dry), Yield = 532 mg (51%)

Figure 85; Summary of the 3.5 mmol scale DIBAL-H reductions of ROY to 12, including yields.

3.4 Intramolecular Environment of ROY Derivatives

During the analysis of the intermolecular environments of the 12 unique ROY polymorphs in Chapter 2 a number of trends were observed. Utilising the Mogul Geometry Check (MGC) feature of Mercury it was identified that the torsion between the two aromatic rings (herein referred to as θ_{thio}) was the main differentiating feature of the ROY polymorphs. Investigation of the distribution of this θ_{thio} value in structures containing similar functionality from within the Cambridge Structural Database (CSD), along with some refinement of the MGC dataset, revealed noticeable differences in trends of θ_{thio} values for structures which are Matched-Molecular Paris (MMPs) of ROY, vs those structures displaying larger structural differences.

In this section MGC is used to evaluate the intramolecular environments of the 12 ROY derivatives synthesised for section 3.1, to compare the distribution of θ_{thio} values observed for the derivatives to the dataset of unique structures containing similar functionality to ROY discussed in section 2.1. Any intramolecular features of the ROY derivatives which are highlighted as *unusual* by MGC are also discussed.

Table 7 contains a summary of selected crystallographic information of crystals of the 12 ROY derivatives synthesised for this work. All crystallographic work was performed at the University of Leeds by the researcher. Data collection was performed by the researcher under the supervision of Christopher Pask. Data reduction and structural optimisation was performed by the researcher and Christopher Pask.

Table 7; Selected crystallographic information for the 12 ROY derivatives synthesised as part of this work.

Form	CSD Deposition Number	Space Group	Z, Z'	θ_{thio} / °
1 ON	2470684	P-1	4, 2	42.8 (Molecule A), 42.2 (Molecule B)
2 R	2471233	P-1	2, 1	7.6
3 RN	2471239	P $2_1/c$	4, 1	53.4
4 R	2471236	P-1	2, 1	28.9
5 ON	2471232	Pbca	8, 1	113.3
6 RN	2473203	P $2_1/c$	n/a	24.8
7 Y	2471234	P-1	2, 1	125.2
8 RN	2471231	P-1	2, 1	32.4
9 DR	2471235	P-1	2, 1	0.4
10 R	2471238	P-1	4, 2	0.5
11 C	2471237	P $2_1/c$	4, 1	16.0
12 RN	2471978	P 2_1 2_1 2_1	4, 1	1.3
12 ON	2471979	P $2_1/n$	4, 1	1.7

3.4.1 Derivatives 1 - 4 Intramolecular Environment

Derivatives **1** – **4** constitute the ROY derivatives in which the four aromatic protons of the nitrobenzene ring (found at positions 3', 4', 5' and 6') of the ROY scaffold have been systematically replaced with methyl (-CH_3) groups. Figure 86; Chemical structures of ROY derivatives 1 – 4. Structural differences to the ROY scaffold are highlighted in red. contains chemical structures for derivatives **1** – **4**.

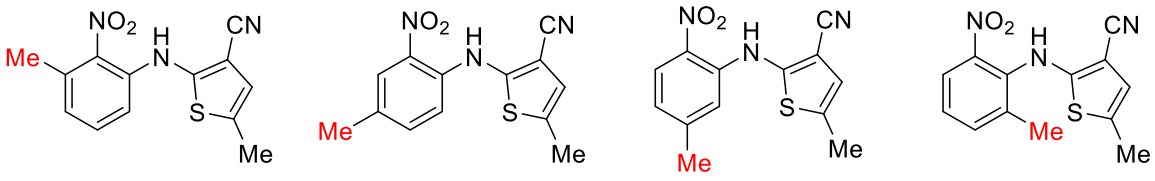


Figure 86; Chemical structures of ROY derivatives 1 – 4. Structural differences to the ROY scaffold are highlighted in red. Full data collection, refinement and optimisation were performed by the researcher

Derivatives **1** – **4** were all crystallised from an excess of EtOAc. Single-Crystal X-Ray Diffractometry (SC-XRD) was performed using a SuperNova X-Ray Diffractometer at 100 K using a Cu-β radiation source. Crystallographic data for derivatives **1** – **4** is summarised in Table 8.

Mogul Geometry Check (MGC) was used to analyse the intramolecular environment of the crystals of derivatives **1** – **4**, and no features labelled as *unusual* were identified. MGC was also used to investigate the θ_{thio} torsions in derivatives **1** – **4**, resulting in the following values: **1** = 42.8° (Molecule A) + 42.2° (Molecule B), **2** = 7.6°, **3** = 53.4° and **4** = 28.9°.

In section 2.1 MGC utilised the ROY scaffold to construct a dataset of instances of the θ_{thio} torsion from across the CSD, which was subsequently refined to contain only unique structures, known as the **Unique Structures Set**. The Unique Structures Set allows for comparison of the θ_{thio} value observed in crystals of derivatives **1** – **4** to those of structures containing the same torsion from within the CSD. Figure 87 contains a histogram describing the distribution of θ_{thio} values in the unique structures set, along with the θ_{thio} values observed n crystals of derivatives **1** – **4** highlighted.

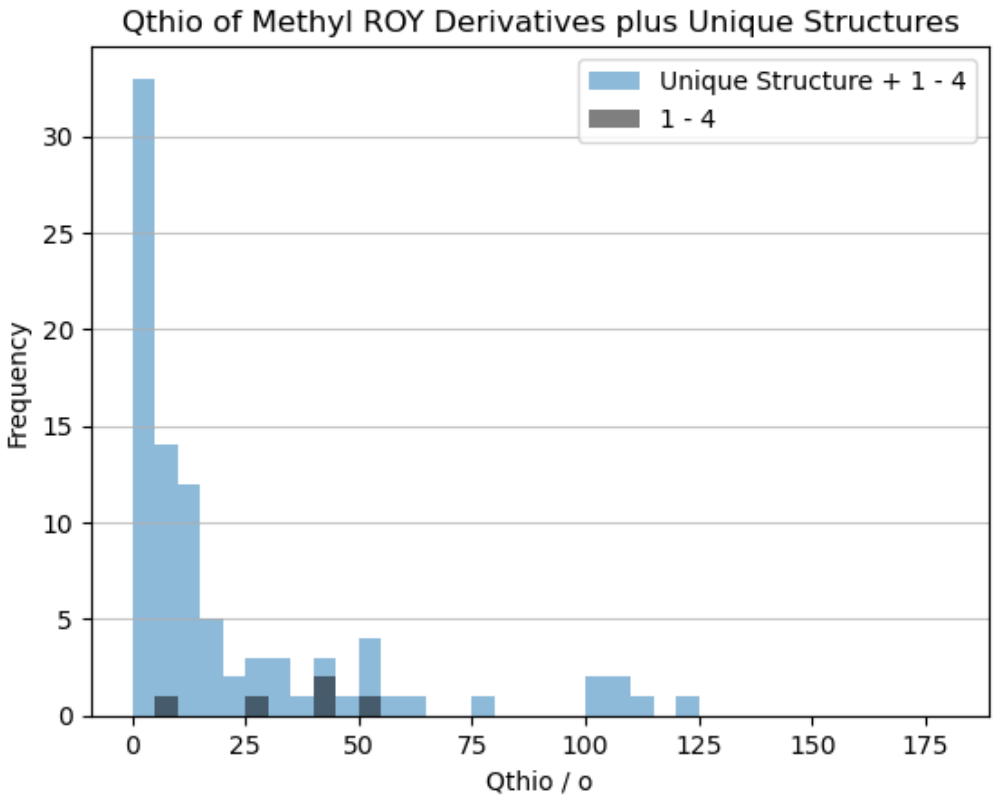


Figure 87; a histogram describing the distribution of θ_{thio} values of crystals of ROY derivatives 1 – 4 in the context of structures from within the Unique Structures Set.

Analysis of the Unique Structures Set in section 2.1 found that the θ_{thio} values of ROY polymorphs and MMPs of the ROY scaffold followed different trends to structures with a higher degree of structural difference to the ROY scaffold. Highlighting the MMPs of ROY within the Unique Structures Set reveals that molecules in the crystals of derivatives **1** – **4** adopt θ_{thio} values more similar to those observed for MMPs of the ROY scaffold. Figure 88 contains a histogram describing the distribution of θ_{thio} values in the Unique Structures Set, highlighting derivatives **1** – **4**, MMPs of ROY, and crystal of molecules of higher structural differences to the ROY scaffold.

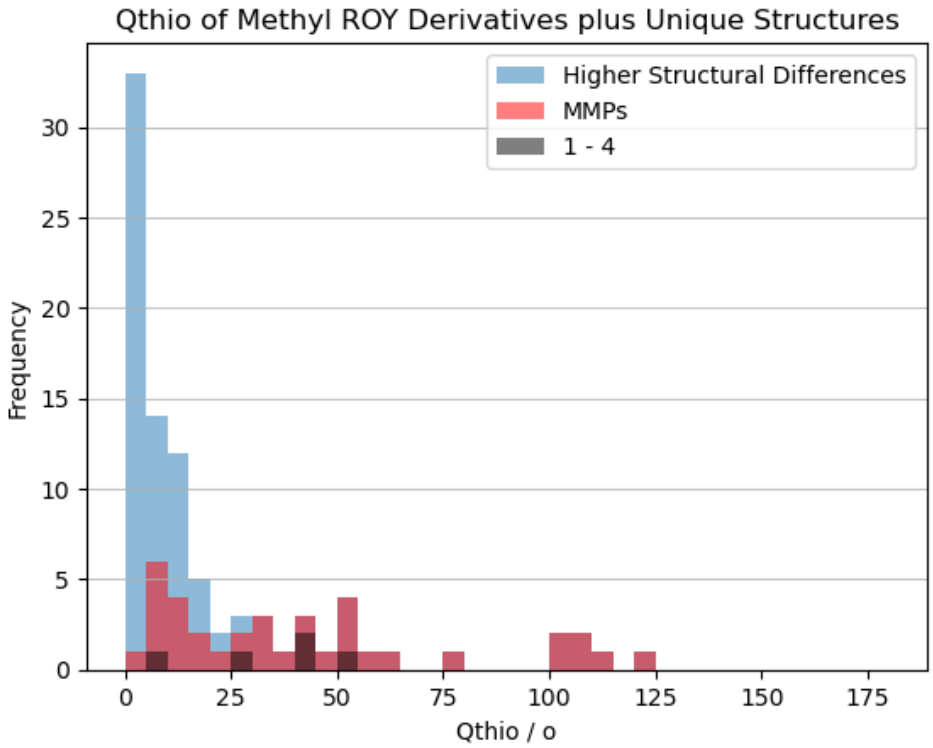


Figure 88; a histogram describing the distribution of θ_{thio} values of crystals of ROY derivatives 1 – 4 in the context of structures from within the Unique Structures Set. MMPs of ROY are highlighted in red, derivatives 1 – 4 are highlighted in black, crystals of higher structural differences are highlighted in blue.

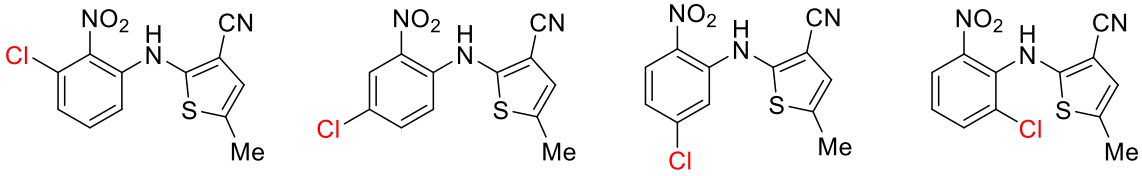
Table 8; Selected crystallographic data for crystals of ROY derivatives 1 – 4. All crystals grown from an excess of EtOAc.

Derivative	**1**	**2**	**3**	**4**
Space Group	P -1 (2)	P -1 (2)	P $2_1/c$ (14)	P -1 (2)
a / A	7.3468(8)	7.7317(10)	8.0878(2)	7.3677(5)
b / A	13.0093(11)	7.9723(11)	13.6488(3)	8.2982(6)
c / A	13.5105(10)	11.4719(13)	11.9221(2)	11.2582(6)
a / °	89.873(7)	82.839(10)	90	88.854(5)
b / °	89.913(7)	76.795(10)	107.143(2)	81.175(5)
g / °	86.767(8)	65.066(13)	90	65.599(7)
Volume / A^3	1289.23	623.937	1257.6	618.688
Z’	2	1	1	1
R^2 / %	4.96	3.83	3.28	3.45

Crystal Habit	Orange Needle	Red Prism	Red Needle	Red Prism

3.4.2 Derivatives 5 – 8 Intramolecular Environment

Derivatives **5** – **8** constitute the ROY derivatives in which the four aromatic protons of the nitrobenzene ring (found at positions 3', 4', 5' and 6') of the ROY scaffold have been systematically replaced with chloro (-Cl) groups. Figure 89 contains chemical structures for derivatives **5** – **8**.



5 **6** **7** **8**

Figure 89; Chemical structures of ROY derivatives 5 – 8. Structural differences to the ROY scaffold are highlighted in red. Full data collection, refinement and optimisation were performed by the researcher

Derivatives **5** – **8** were all crystallised from an excess of EtOAc. Single-Crystal X-Ray Diffractometry (SC-XRD) was performed using a SuperNova X-Ray Diffractometer at 100 K using a Cu-β radiation source. Crystallographic data for derivatives **5** – **8** is summarised in **Table 9**.

MGC was used to analyse the intramolecular environments of the molecules in crystals of derivatives **5** – **8**, which identified the C7-C6-N2-O2 torsion in derivative **5** as being unusual (40.86°). This torsion refers to the twisting of the nitro group relative to the attached benzene ring, and is highlighted in **Figure 90** for derivative **5**.

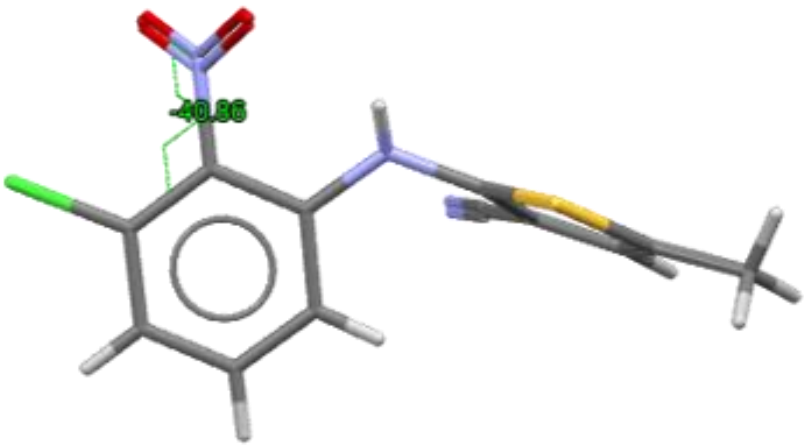


Figure 90; simulation of molecules within derivative 5, with the θ_{nitro} torsion highlighted.

This twisting of the nitro group relative to the attached benzene ring (herein referred to as θ_{nitro}) is also observed for derivative **6** – **8**, but only the θ_{nitro} value for derivative **5** is highlighted as unusual by MGC. This θ_{nitro} torsion is not observed to any significant degree for any of the 12 ROY polymorphs, nor derivative **1** – **4**, and therefore is assumed to be influenced by the introduction of the chlorine atom to the benzene ring. The unusual θ_{nitro} torsion present in derivative **5** suggests that other polymorphs of 5 may be possible.

Analysis of the θ_{thio} values for derivatives **5** – **8** demonstrates that the θ_{thio} torsion of the ROY scaffold and MMPs of ROY (such as derivatives **5** - **8**) behave more similarly than non-MMPs of the ROY scaffold within the **Unique Structures Set**. **Figure 91** contains a histogram describing the distribution of θ_{thio} values of MMPs of ROY, non-MMPs of ROY and derivatives **5** – **8** from within the **Unique Structures Set**.

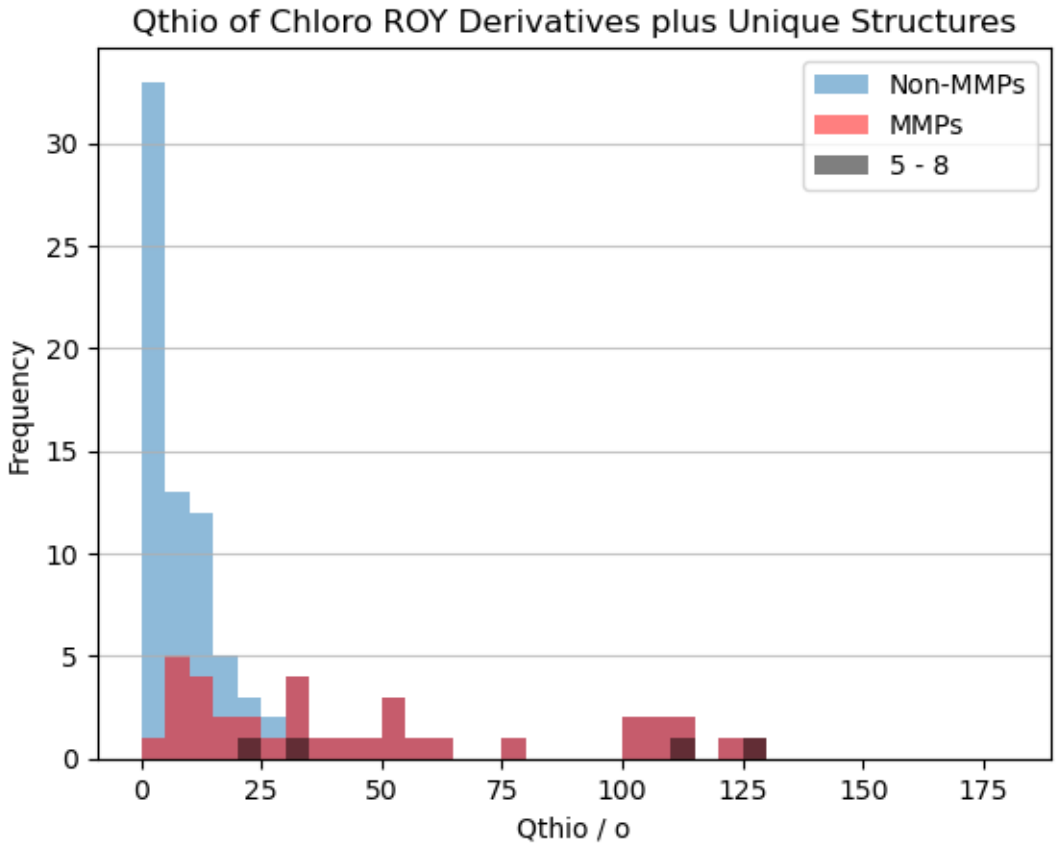


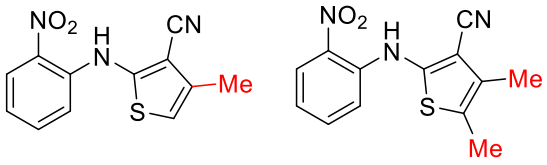
Figure 91; a histogram describing the distribution of θ_{thio} values of crystals of ROY derivatives 5 – 8 in the context of structures from within the Unique Structures Set. MMPs of ROY are highlighted in red, derivatives 5 – 8 are highlighted in black, crystals of higher structural differences are highlighted in blue.

Table 9; Selected crystallographic data for crystals of ROY derivatives 5 – 8. All crystals grown from an excess of EtOAc.

Derivative	**5**	**6**	**7**	**8**
Space Group	P b c a (61)	P 2_1/c (14)	P -1 (2)	P -1 (2)
a / A	16.2513(12)	3.81240(10)	7.2709(4)	7.4288(3)
b / A	7.9098(4)	14.6291(4)	8.2768(4)	8.3450(3)
c / A	19.4915(10)	20.8443(7)	10.8833(7)	11.3254(4)
α / ^{o}	90	90	75.393(5)	88.719(3)
β / ^{o}	90	89.528(3)	87.833(4)	80.096(3)
γ / ^{o}	90	90	82.266(4)	67.608(3)
Volume / A^3	2505.53	1162.49	628.015	638.768
Z'	1	0.75	1	1
R^2 / %	4.84	8.3	3.07	4.06
Crystal Habit	Orange Needle	Red Needle	Yellow Prism	Red Needle

3.4.3 Derivatives 9 and 10 Intramolecular Environment

Derivatives **9** and **10** constitute the ROY derivatives in which the 4 and 5 positions, Figure 92 contains chemical structures for derivatives **9** and **10**.



9 **10**

Figure 92; Chemical structures of ROY derivatives 9 and 10. Structural differences to the ROY scaffold are highlighted in red.

Crystals of derivatives **9** and **10** were crystallised from an excess of EtOAc. Single-Crystal X-Ray Diffractometry (SC-XRD) was performed using a SuperNova X-Ray Diffractometer at 100 K using a Cu-β radiation source. Crystallographic data for derivatives **9** and **10** are summarised in **Table 10; Selected crystallographic data for crystals of ROY derivatives 9 and 10. All crystals grown from an excess of EtOAc.**Table 10.

MGC analysis of the structures of derivatives **9** and **10** highlighted no unusual intramolecular features present in either crystal structure. Analysis of the θ_{thio} torsion in derivatives **9** and **10** revealed the θ_{thio} value to be 0.35° for derivative **9**, and 0.52o **10**. Whilst these θ_{thio} values are not unusual amongst the Unique Structures Set, there are few MMPs of ROY which adopt similar torsions. In the MMPs of ROY set, there is only one example of a θ_{thio} value in the range $0° < \theta_{thio} < 5°$, which belongs to molecule B of the QAXBOK crystal. Image of the θ_{thio} torsions present in derivative **9**, **10** and QAXBOK can be found in Figure 93.

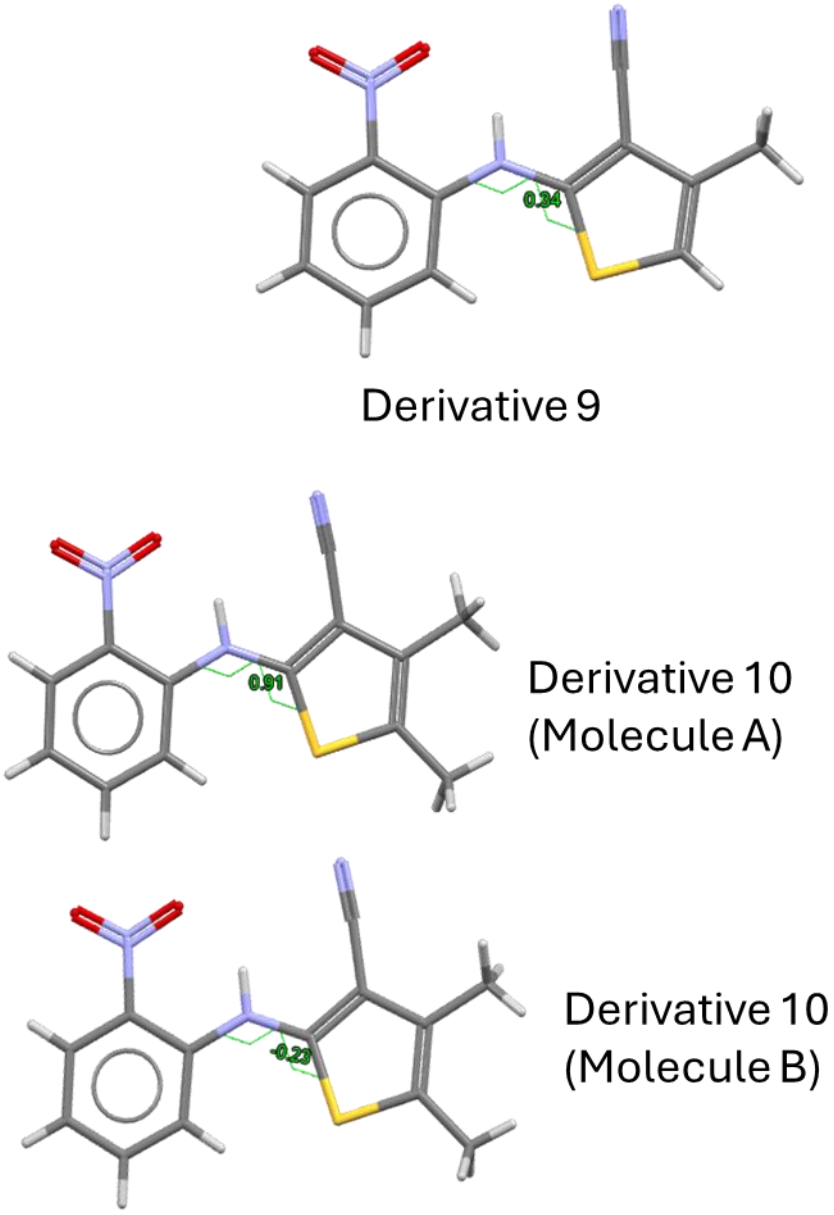


Figure 93; (Top) central molecule within derivative 9 with the θ_{thio} torsion highlighted. (Bottom) θ_{thio} torsion highlighted in molecule A and B in derivative 10.

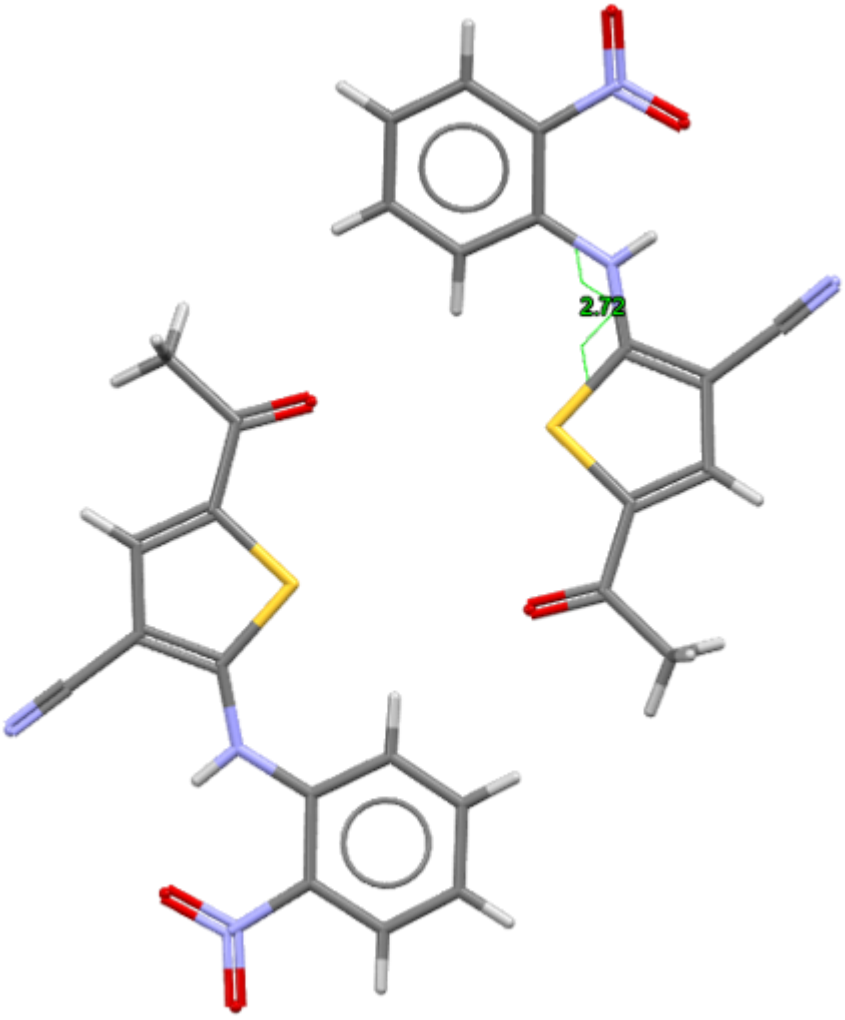


Figure 94; Molecule B of within the asymmetric unit of the 5-acetly-ROY form QAXBOK with θ_{thio} highlighted.

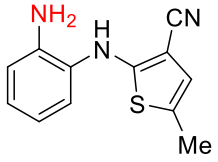
The θ_{thio} values observed for derivatives 9 and 10 are lower than any θ_{thio} value observed for MMPs of ROY within the Unique Structures Set, which could be considered unusual torsional behaviour.

Table 10; Selected crystallographic data for crystals of ROY derivatives 9 and 10. All crystals grown from an excess of EtOAc.

Derivative	**9 DR**	**10 R**
Space Group	P -1 (2)	P 2_1/c (14)
a / A	7.4689(5)	6.7484(5)
b / A	8.0675(7)	12.2195(9)
c / A	11.2854(6)	15.3353(12)
α / °	104.806(6)	90
β / °	96.361(5)	99.014(10)
γ / °	115.625(8)	90
Cell Volume / A^3	573.343	1248.96
Z'	1	1
R^2 / %	3.67	4.92
Crystal Habit	Dark Red Prism	Red Prism

3.4.4 Derivative 11 Intramolecular Environment

Derivative **11** is a derivative of the ROY scaffold in which the nitro group attached to the benzene ring is reduced to the corresponding primary amine. Derivative **11** is synthesised commercially as a precursor to the dug olanzapine, and as such data for a crystal form of **11** are available, including an entry to the CSD (refcode = XABXEF). Figure 95 contains a chemical structure for derivative **11**.



11

Figure 95; Chemical structure for ROY derivative 11. Structural differences to the ROY scaffold are highlighted in red. Full data collection, refinement and optimisation were performed by the researcher

Crystals of derivative **11** were crystallised from an excess of EtOAc. Single-Crystal X-Ray Diffractometry (SC-XRD) was performed using a SuperNova X-

Ray Diffractometer at 100 K using a Cu-β radiation source.. Crystallographic data for derivative **11** and CSD entry XABXEF are summarised in Table 11.

Comparison of the crystallographic data from the Crystal Information Files (CIFs) for derivative **11** and XABXEF reveals similarity between the two datasets. Furthermore, analysis of the θ_{thio} torsions in derivative **11** and XABXEF shows the two values are nearly identical (**11** = 16, XABXEF = 16).

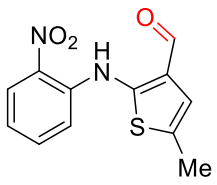
Whilst it would require further data to confirm this, such as direct comparison of PXRD patterns or thermal analysis of the crystals, the similarities observed in crystallographic data and intramolecular features of derivative **11** and XABXEF suggest that the two are identical crystal forms.

Table 11; Selected crystallographic data for crystal of derivative 11. Crystallographic data for XABXEF is provided for reference.

Derivative	**11**	XABXEF
Space Group	P 2_1/c (14)	P 2_1/c (14)
a / A	12.9251(7)	12.9188(7)
b / A	8.2256(3)	8.2257(4)
c / A	11.2061(5)	11.2046(6)
α / °	90	90
β / °	109.763(6)	109.784(3)
γ / °	90	90
Cell Volume / A^3	1121.22	1120.39
Z'	1	1
R^2 / %	4.66	4.35
Crystal Habit	Colourless Prism	N/a

3.4.5 Derivative 12 Intramolecular Environment

Derivative **12** is a derivative of the ROY scaffold in which the nitrile group attached to the thiophene ring is partially reduced to the corresponding aldehyde. **Figure 96** contains a chemical structure for derivative **12**.



12

Figure 96; Chemical structure for ROY derivative 12. Structural differences to the ROY scaffold are highlighted in red. Full data collection, refinement and optimisation were performed by the researcher

In experiments the performed in an attempt to isolate crystals of sufficient quality for SC-XRD, multiple crystal forms of derivative **12** were discovered. The first, a red needle (**12 RN**), was isolated via crystallisation from excess EtOAc, and the second, an orange needle (**12 ON**) was isolated via crystallisation from excess chloroform. Single-Crystal X-Ray Diffractometry (SC-XRD) was performed using a SuperNova X-Ray Diffractometer at 100 K using a Cu-β radiation source. Crystallographic data for derivative **12 RN** and derivative **12 ON** are summarised in Table 12.

MGC analysis of the intramolecular environments presented by 12 RN and **12 ON** highlighted no intramolecular features as unusual. Analysis of the θ_{thio} values present in **12 RN** and **12 ON** identifies θ_{thio} values of 1.33° and 1.67°, respectively. Similar to the derivatives **9** and **10** discussed in section 3.2.3, the θ_{thio} values presented by **12 RN** and **12 ON** would not be highlighted as unusual amongst the Unique Structures Set derived from the CSD, despite displaying θ_{thio} values lower than any MMP of ROY within the Unique Structures Set. However, by splitting the Unique Structures Set into MMPs and non-MMPs of ROY it is revealed that there is only one MMP structure found in the range in which **12 RN** and **12 ON** are found (QAXBOK, see Figure 94)

It is interesting to note that derivatives **9**, **10**, **12 RN** and **12 ON** all display θ_{thio} values which are very low for MMPs of ROY, and all feature modifications to the ROY structure based on the aminothiophene ring. In the forms of derivative **12**, the planarity of molecules within the crystal structure can be explained by the influence of the six-membered ring intramolecular H-bond formed between the bridging amine proton and the aldehyde oxygen present in **12**.

Another interesting note is the similarity of θ_{thio} values observed for **12 RN** and **12 ON**, with a difference of only 0.34°, this is a significantly smaller difference in θ_{thio} value than is observed between any two of the unique ROY polymorphs within the CSD. This could suggest that the apparent restriction of torsional flexibility present in **12** has little impact on the accessibility of multiple polymorphic forms of **12**.

Table 12; selected crystallographic data for crystals of derivative 12 RN, grown from excess EtOAc, and 12 ON, grown from excess chloroform.

Derivative	**12 RN**	**12 ON**
Space Group	P 2_1 2_1 2_1 (19)	P 2_1/n (14)
a / A	3.8649(2)	6.8092(3)
b / A	16.6048(11)	12.5856(4)
c / A	17.5166(10)	13.7657(5)
α / ^{o}	90	90
β / ^{o}	90	102.310(4)
γ / ^{o}	90	90
Cell Volume / A^3	1124.14	1152.57
Z'	1	1
R^2 / %	5.71	5.82
Crystal Habit	Red Needle	Orange Needle

3.5 Intermolecular Environment of ROY Derivatives

In Chapter 2.2 the intermolecular environments of the 12 ROY polymorphs within the CSD were explored and evaluated by considering the intermolecular contacts observed in each form in the context of Full Interaction Maps (FIMs) calculated for central molecules in each polymorph. This evaluation was used to provide a rationalisation for the relative polymorphic stability observed between the 12 ROY polymorphs, along with identifying key functional groups within the ROY scaffold which play a role in determining the overall stability of a given ROY form. In descending order of their influence on polymorphic stability, these functional groups are: the nitrile group, the nitro group, aromatic protons and methyl groups.

In Chapter 3.1, the observations taken from the analysis of the intermolecular environments of the 12 ROY polymorphs within the CSD were then used to guide the design a set of 12 functional derivatives of the ROY scaffold, with the modifications made to the ROY scaffold designed to probe the influence of the nitrile, nitro, aromatic proton and methyl groups on polymorphic stability of these ROY derivatives.

In this section, the goal is to use FIMs in conjunction with an evaluation of the intermolecular contacts observed in the crystal form of each ROY derivative to

explore the intermolecular environments of the 12 ROY derivatives synthesised in Chapter 3.1. The goal of this analysis is to evaluate the role and influence of the modifications made to the ROY scaffold on the intermolecular environment of derivatives **1** – **12**, to explore what (if any) effect these modifications have on the remaining influential groups within the scaffold, and finally to provide a base for assessment of the likely polymorphic profile of derivatives **1** – **12**.

3.5.1 Derivatives 1 – 4 Intermolecular Environment

Derivatives **1** – **4** constitute ROY derivatives in which the C-H bonds observed at the 3' – 6' position of the nitrobenzene ring have been systematically replaced with C-CH_3 bonds.

Orange needles of Derivative 1 were isolated via slow crystallisation from an excess of EtOAc. **1 ON** is the only derivative discussed in this section with Z' = 2, and one of only three Z' = 2 structures within the library of forms of ROY derivatives produced for this work. The θ_{thio} values for molecule A and molecule B within the crystal lattice of **1** are 42.78° and 42.19°, respectively, suggesting that the only major difference in the intermolecular environments observed by each molecule is present in the intermolecular contacts formed by each molecule.

As was observed for the **R18** (QAXMEH57) Form of ROY in Section 2.2.1.2, intermolecular contacts in structures with Z' = 2 are complimentary, with functional groups of molecule A forming intermolecular contacts with functional groups of molecule B and vice versa. This is with one exception, which is a single functional group present in one of the molecules which forms intermolecular contacts which are not reciprocated by the other molecule in the asymmetric unit, exemplified by the different nitrile nitrogen environments observed by molecule A and molecule B in the **R18** (QAXMEH57) Form of ROY. As such, for **1 ON**, FIMs and intermolecular contacts are only calculated for molecule B within the asymmetric unit, due to the nitro oxygen of molecule A of derivative showing a different environment to the one observed around the nitro oxygen of molecule A.

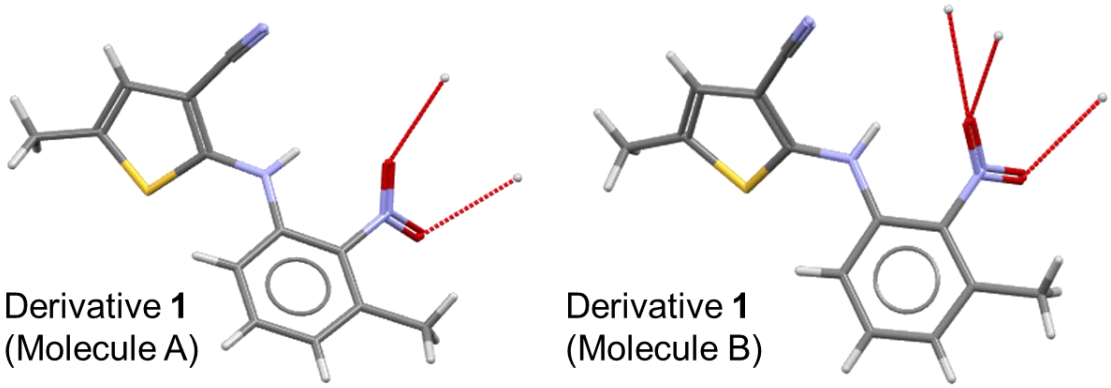


Figure 97; (left) molecule A within the asymmetric unit of derivative 1, with the intermolecular short contacts formed within the crystal lattice highlighted in red. (Right) molecule B within the asymmetric unit of derivative 1, with the intermolecular short contacts formed within the crystal lattice highlighted in red.

FIMs and intermolecular contacts calculated for molecule B within the crystal lattice of **1 ON** can be found in Figure 98. In order for direct evaluation of the change in intermolecular environment introduced by the structural changes to the ROY scaffold present in the ROY derivatives, the probe groups used to generate the FIMs hotspots for Molecule B of **1 ON** are the same as those used to analyse the intermolecular environments for the ROY polymorphs in Section 2.2.1.

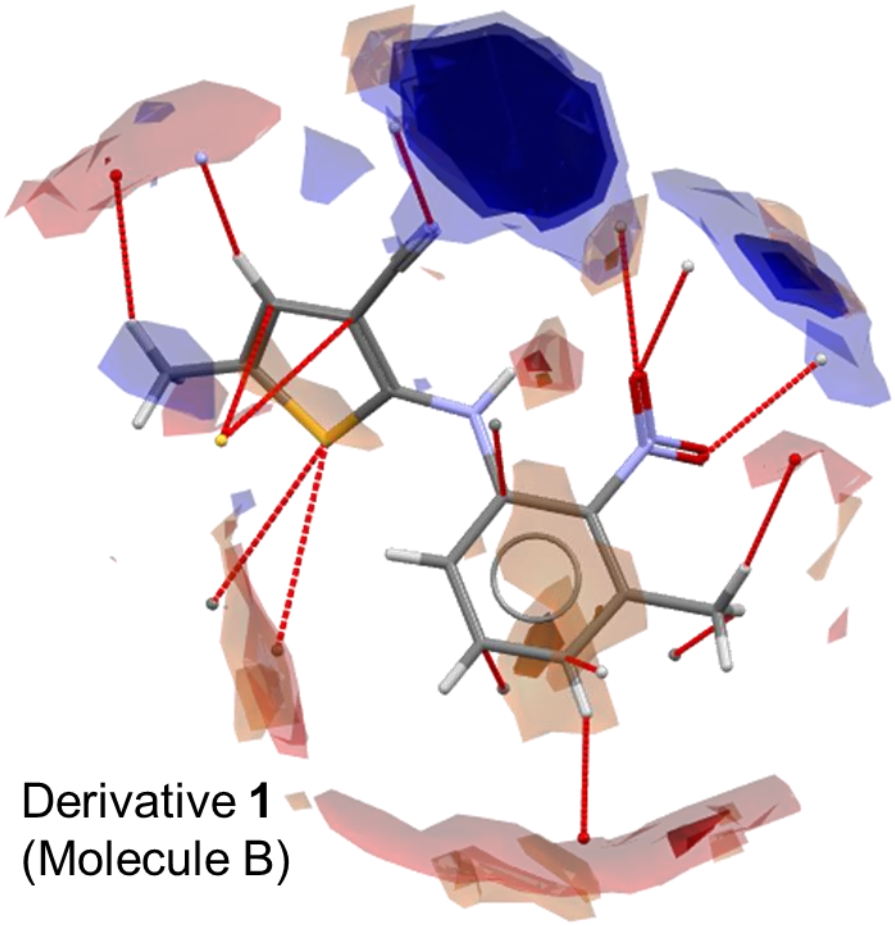


Figure 98; FIMs and intermolecular short contacts calculated for molecule A within the asymmetric unit of derivative 1, using uncharged NH hydrogen (blue), carbonyl oxygen (red) and aromatic CH hydrogen (brown) probes used to calculate FIMs.

The first to thing to note for molecule B of **1 ON** is the impact of the additional C-CH_3 group at the 3' position of the nitrobenzene ring. The methyl group replaces the C-H group present at that position in the **ON** (QAXMEH) Form of ROY (Figure 46) resulting in a reduction in the area covered by the light red (carbonyl oxygen) hotspot observed around the plane of the nitrobenzene ring. Despite this, protons attached to the methyl group are involved in two separate intermolecular contacts. One of these contacts is formed with the ntiro O3 oxygen of molecule A, whilst the other is formed with the nitrobenzene C8 carbon of a neighbouring molecule B. An image of the intermolecular environment around the 3' methyl group, including its intermolecular contacts, can be found in Figure 97. The H26 proton – O3 oxygen contact is located in the red (carbonyl oxygen) hotspot associated with the 3' methyl group, and could be considered favourable for the methyl group, however the interaction with C8 resembles a π-type interaction; a class of interactions observed prominently amongst metastable ROY polymorphs.

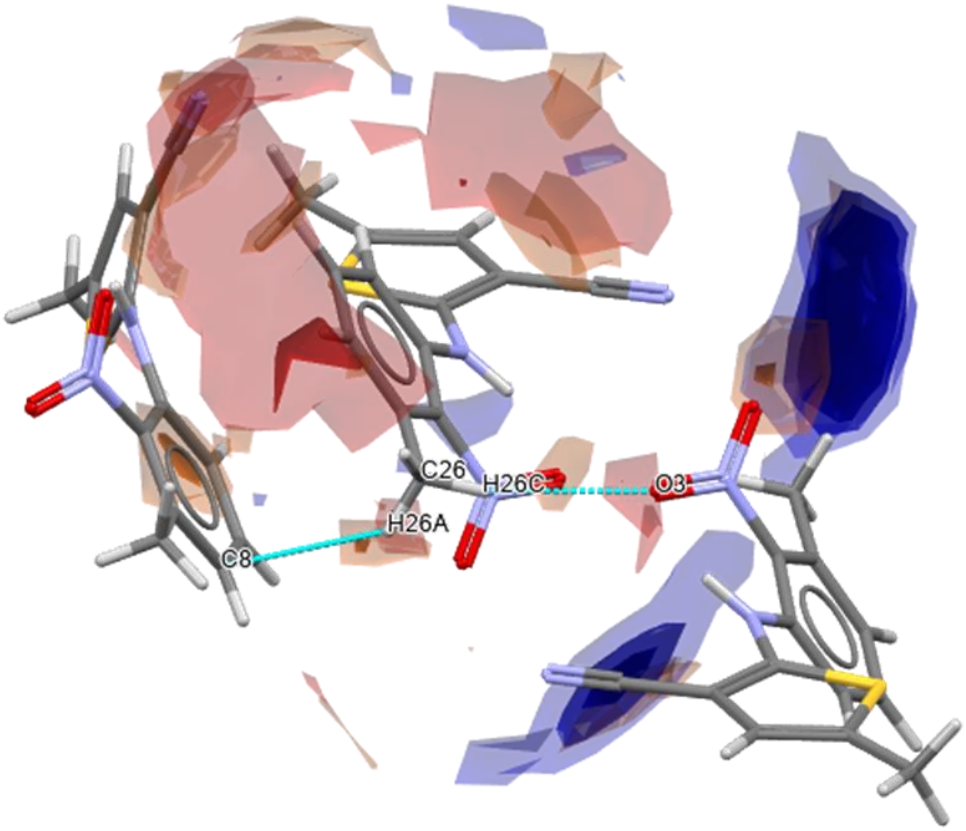


Figure 99; intermolecular environment around the C26 methyl group located at position 3' of the thiophene ring of molecule B of derivative 1.

Analysis of the stable and metastable polymorphs from Chapter 2.2 highlighted the nitrile group as playing an important role in the stabilisation of ROY Forms, with a favourable nitrile environment being indicative of more stable Forms. The nitrile group in 1 ON forms a contact with the thiophene H16 proton at position 4 of the thiophene ring, as visualised in Figure 98. This contact falls at the edge of the dark blue (uncharged NH hydrogen) hotspot associated with the nitrile nitrogen, and also at the edge of the light red (carbonyl oxygen) hotspot associated with the thiophene H16 proton, suggesting an overall favourable interaction.

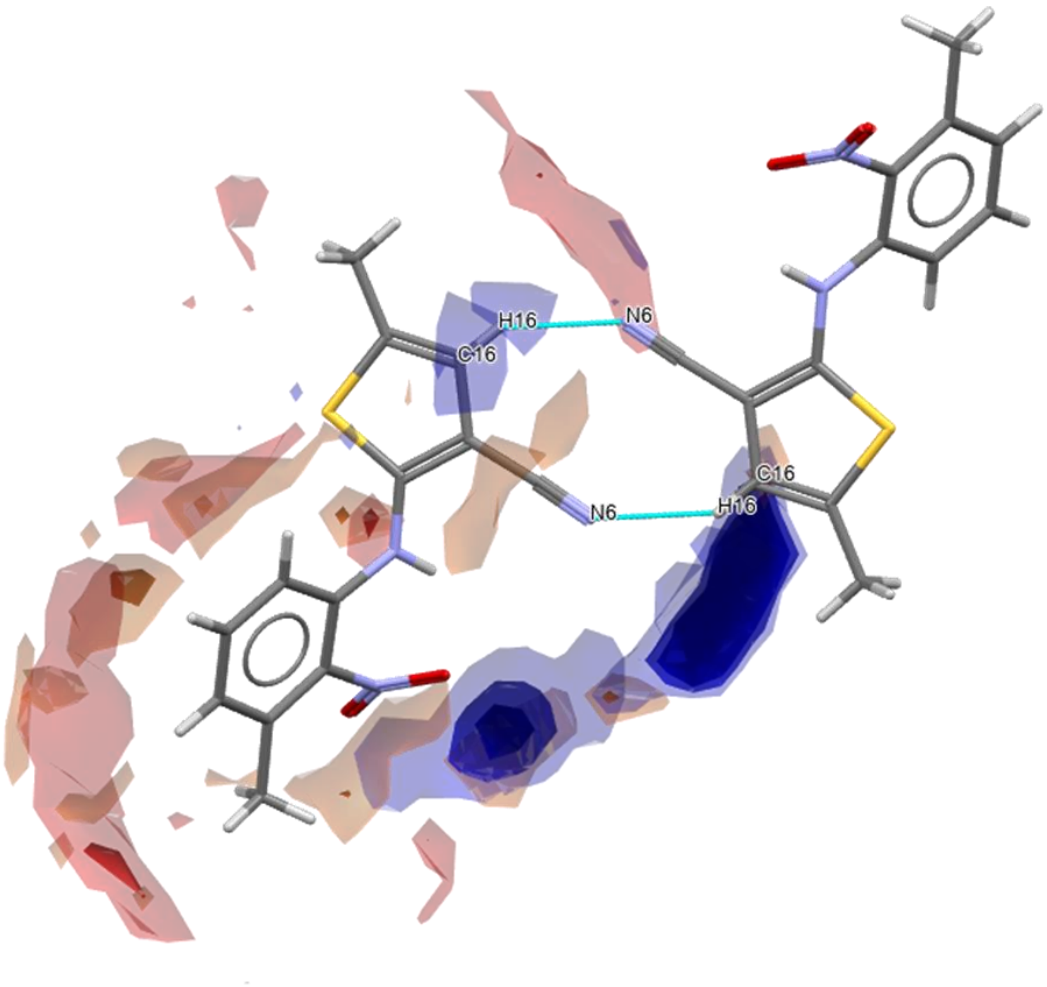


Figure 100; FIMs calculated for molecule B of derivative 1, along with the intermolecular contact between the nitrile N6 nitrogen and the thiophene H16 proton of a neighbouring molecule B from within the crystal lattice of 1.

The nitro group was highlighted in Chapter 2.2 as playing a key role in the stabilisation of ROY polymorphs, in a similar yet less prominent fashion to the effect of the nitrile group on polymorphic stability. As such, a favourable nitro environment is considered an indicator of more stable polymorphic forms of ROY. The nitro environment is also the only major difference observed between the intermolecular environments of molecule A and molecule B of 1. A visualisation of the nitrile environment of molecule B of 1 can be found in Figure 99.

The nitro O4 atom is involved in a single intermolecular contact with the H8 proton at the 4' position of the nitrobenzene ring of a neighbouring molecule A, which is located near the edge of the blue (uncharged NH hydrogen) hotspot associated with the nitro group, which can be considered a reasonably favourable intermolecular contact.

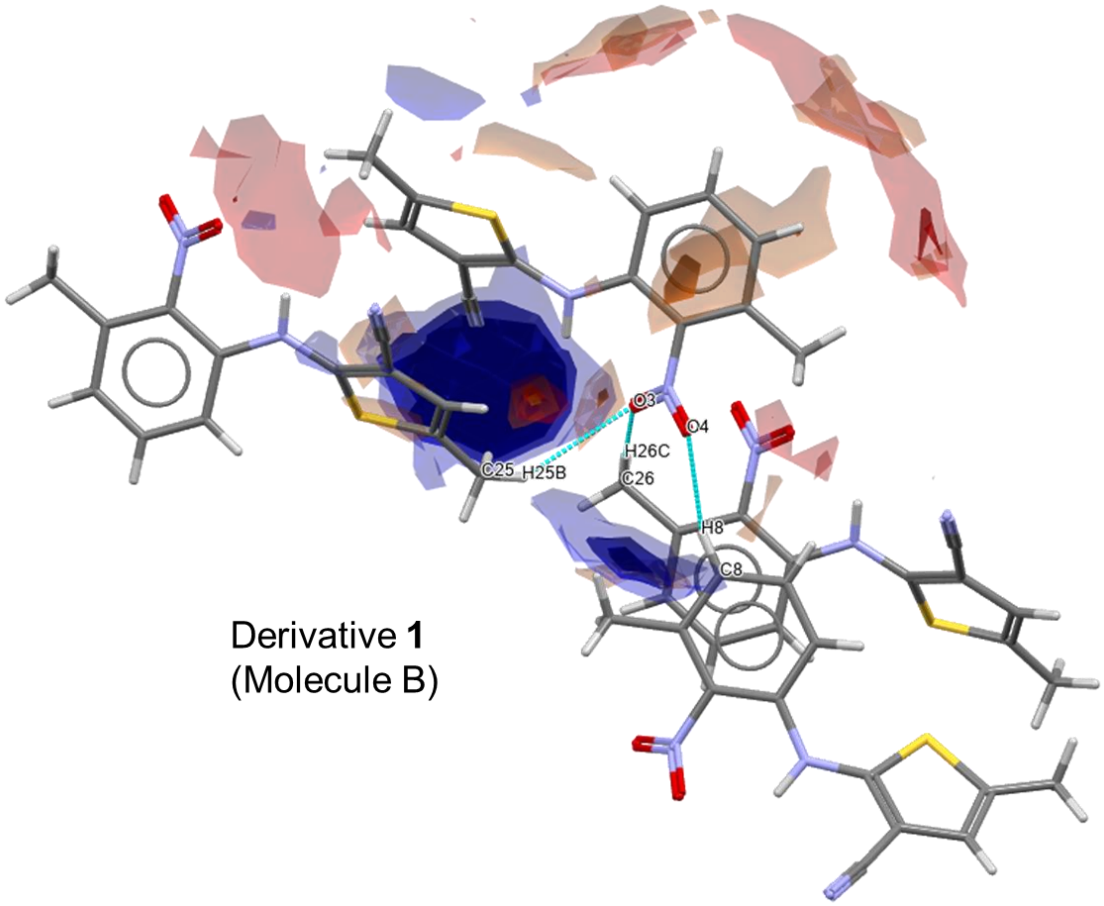


Figure 101; a visualisation of the nitro environment observed for molecule B of 1 ON, highlighting the contact between the nitro O4 and nitrobenzene H8 hydrogen, the nitro O3 atom with the thiophene-bound methyl H25B and nitrobenzene-bound methyl H26C hydrogens.

However, contacts formed by the O3 oxygen seem less favourable. The O3 oxygen forms contacts with two separate methyl-bound protons of separate molecule B's from within the crystal lattice of **1 ON**. One contact is formed with the thiophene-bound methyl proton H25B, whilst the other is formed with the nitrobenzene-bound methyl proton H26C. Both of these contacts around found outside of the blue (carbonyl oxygen) hotspot associated with the nitro oxygen, and, individually, both of these contacts can be considered unfavourable for the nitro group. The combination of the contacts formed by the O3 and O4 atoms suggests an unfavourable environment for the nitro group of molecule B of **1 ON** overall.

Chapter 2.2.1.2 highlighted that significant π-π stacking interactions are observed in the metastable forms of ROY, whilst having much less representation amongst stable forms. **1 ON** displays a number of π-π stacking interactions, which can be found in Figure 100. These contacts are formed between the C5' carbon and C3' proton with the C1' and C4' carbons of neighbouring **1** molecules within the crystal lattice, alternating each time. Whilst these contacts are found within the

brown (CH hydrogen) hotspots associated with the nitrobenzene rings of **1 ON**, the presence and prominence of these contacts is indicative of metastable ROY forms, and hence is considered an indicator of a metastable **1 ON** form.

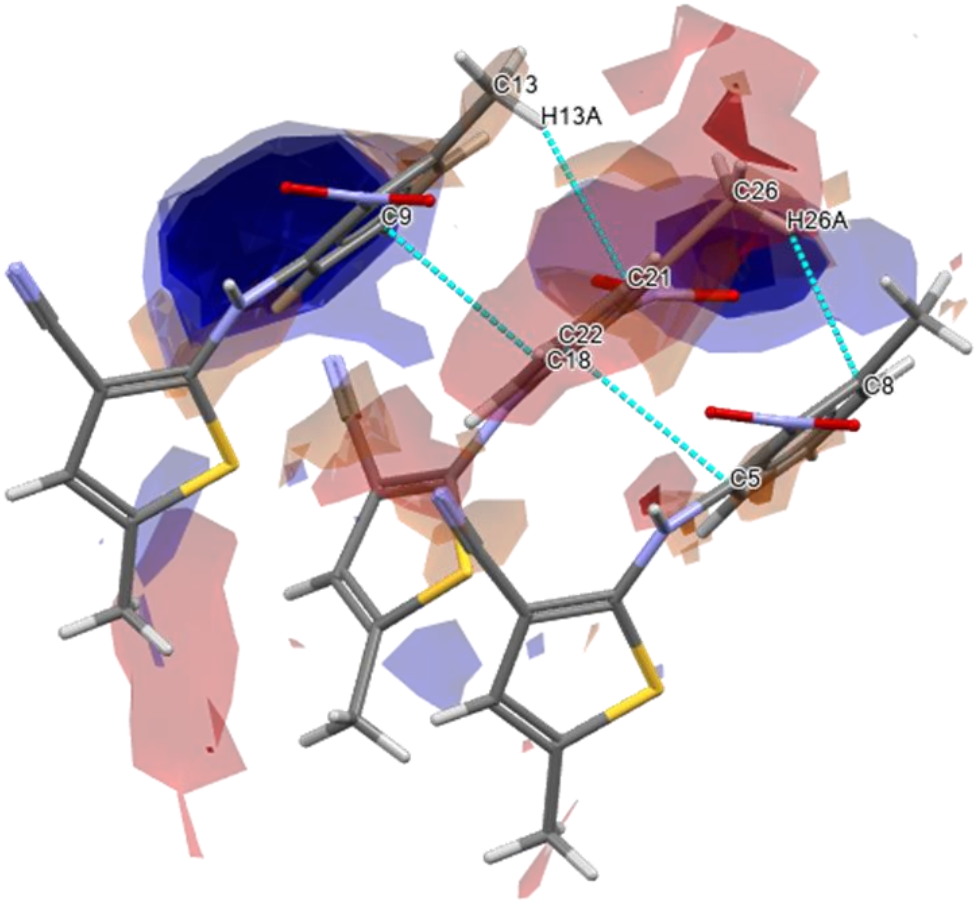


Figure 102; a visualisation of the π-π stacking interactions formed by molecule B of 1 ON.

Overall, there are a number of positive and negative aspects of the intermolecular environment observed in **1 ON**. The favourable nitrile environment present in **1 ON** is a clear indicator of a favourable intermolecular environment, as evidenced by the analysis in Chapter 2.2, however the presence of the unfavourable nitro O3 environment, alongside the prominent π-π stacking interactions, suggests an intermolecular environment that is less than ideal. This suggests that **1 ON** may be amongst the more stable polymorphs of **1**, but that more stable polymorphs could be accessible.

Red prisms of derivative 2 were isolated via slow evaporation from an excess of EtOAc. **2 RP** is unique amongst the forms discussed in this section as it is the only derivative which has been previously reported in the literature (see Figure), for which one structure is found in the CSD, the dark red form OGIDEN. Investigation of the structural information for **2 RP** in section 3.2.1 revealed striking similarities between **2 RP** and OGIDEN, suggesting that they are the

same form. There is no R^2 value associated with the OGIDEN crystal structure, and there are a number of contacts present in **2 RP** which are missing in OGIDEN. As such, the intermolecular environment of the **2 RP** crystal structure is analysed in this section.

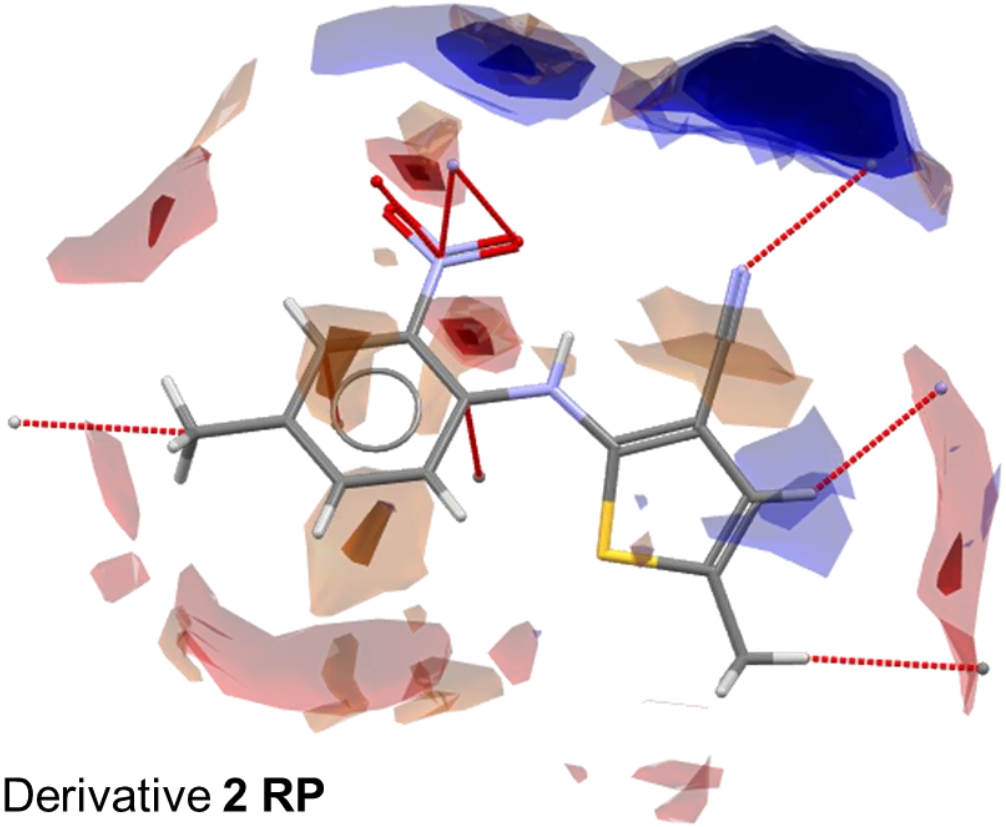


Figure 103; a visualisation of the overall intermolecular environment observed for molecules within the crystal lattice of Derivative 2 RP.

FIMs and short contacts for **2 RP** can be found in Figure 101. At a glance the intermolecular environment of **2 RP** looks favourable, with the majority of intermolecular contacts being located in or near their corresponding FIMs hotspots. The only exception being the methyl-methyl contact formed between the methyl C13 carbon at the 4' position of the nitrobenzene ring and the methyl C12-H12B proton at the 5 position of the thiophene ring. This contact is located within the red (carbonyl oxygen) hotspot associated with the aminothiophene-bound methyl group, but outside of any hotspot associated with the nitrobenzene-bound C13 atom. This suggests a contact which is favourable for the aminothiophene-bound methyl group, but less so for the nitrobenzene-bound methyl group. A visualisation of the methyl-methyl contact present in **2 RP**, alongside FIMs calculated for **2 RP**, can be found in Figure 101.

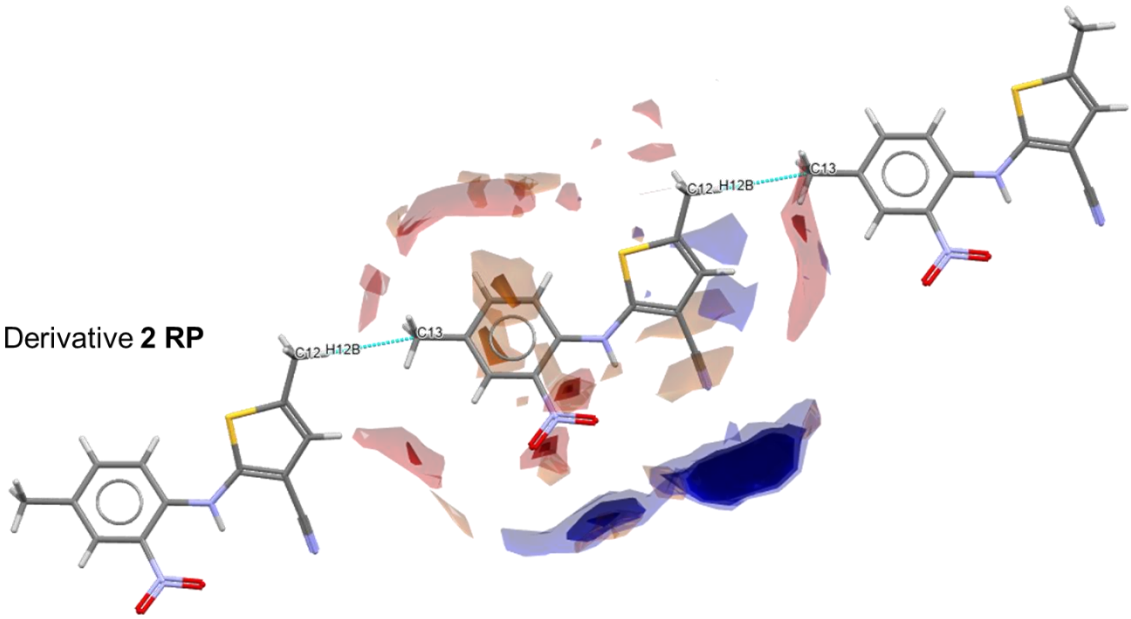


Figure 104; a visualisation of the intermolecular contact formed between the methyl-bound C13 carbon and thiophene-bound methyl H12B hydrogen in 2 RP.

Another feature to note is the NO_2 environment, which displays contacts formed within the dark red (carbonyl oxygen) hotspot located above the plane of the NO_2 group in **2 RP**. These contacts are a self-interaction between two NO_2 groups between neighbouring molecules of **2 RP** within the crystal lattice. The location of the contact within the dark red hotspot associated with the NO_2 group suggests the interaction is favourable. However, the lack of contacts observed in the larger dark blue (uncharged NH hydrogen) hotspot associated with the oxygen atoms of the NO_2 is noteworthy, as it was primarily the occupation of these hotspots which signified favourable NO_2 environments for ROY Forms in Chapter 2.2. Overall, the interactions observed around NO_2 are favourable, but in the absence of contacts around the dark blue hotspots suggests more favourable interactions may be possible. A visualisation of the NO_2 environment present in **2 RP**, alongside FIMs calculated for **2 RP**, can be found in Figure 103.

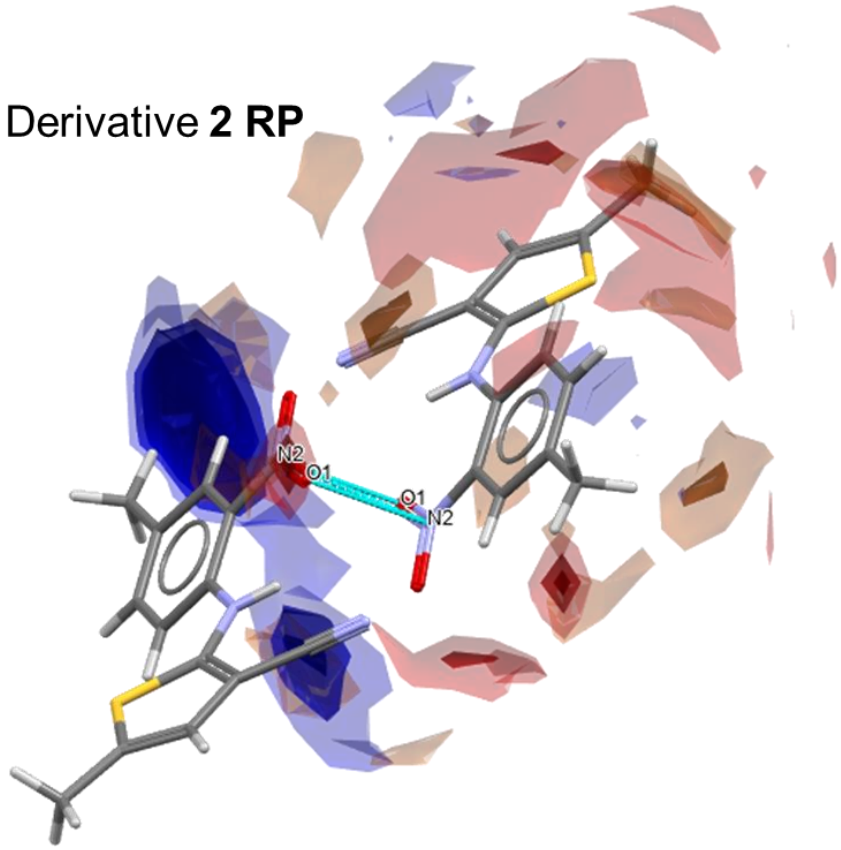


Figure 105; a visualisation of the intermolecular environment observed around the nitro group in 2 RP.

The environment of **2 RP** is visually reminiscent of the environment observed for the **ON** (QAXMEH) form of ROY (Figure 47), one of the most stable ROY forms, and contains a number of favourable interactions. This, alongside the presence of OGIDEN within the CSD, suggesting it to be the most thermodynamically stable of the forms found in the 2001 study, suggests that **2 RP** is among the most stable forms of Derivative **2**.

Red needles of derivative **3** were isolated via slow evaporation from an excess of EtOAc. **3 RN** is a functional analogue of the ROY scaffold in which the C-H bond at the 5'-position of the nitrobenzene ring has been replaced by a C-CH_3 bond. Figure 104 contains a visualisation of the intermolecular contacts formed within the crystal lattice of **3 RN**, along with FIMs calculated for **3 RN**. Immediately there are a number of differences visible between the environment of **3 RN** and the previous structures discussed in this section. The majority of contacts are located outside of their corresponding FIMs hotspots, and there is a complete absence of contacts involving the NO_2 group. There is also a π-π stacking-type interaction associated with the nitrobenzene ring, similar to **1 ON** and **2 RP**.

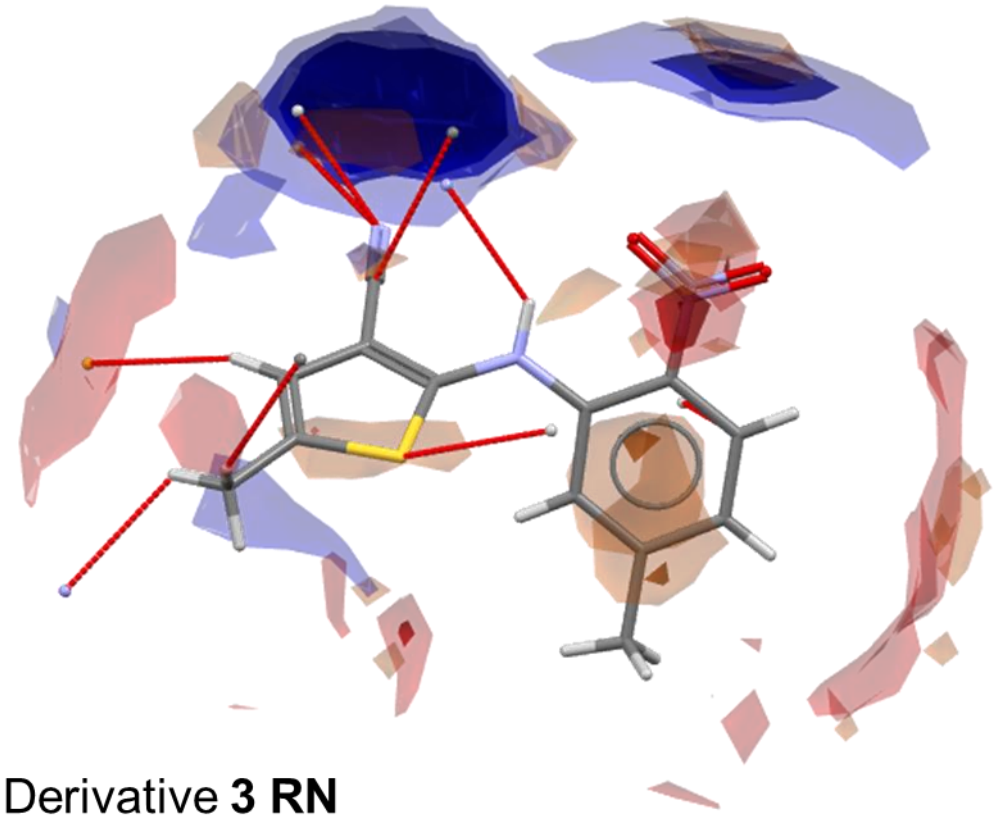


Figure 106; a visualisation of the overall intermolecular environment observed within the crystal lattice of 3 RN.

Another unfavourable feature of the intermolecular environment of **3 RN** is the environment of the nitrile group. The CN group in **3 RN** is involved in three contacts, with the CN carbon forming a self-interaction with a neighbouring CN carbon atom, and the CN nitrogen forming two intermolecular contacts. The first is located within the dark blue (uncharged NH hydrogen) hotspot, which is formed with the thiophene-bound C12-H12B proton, whilst the second is a formal intermolecular H-bond formed with the bridging NH hydrogen. It was found throughout chapter 2 that the presence of CN-HN H-bonds was indicative of more stable polymorphs of ROY, however, the number of contacts associated with the CN group in **3 RN**, along with their position relative to the FIMs hotspots associated with the CN group in **3 RN**, could suggest an overall unfavourable CN environment. A visualisation of the CN environment present in **2 RP**, alongside FIMs calculated for **2 RP**, can be found in Figure 105.

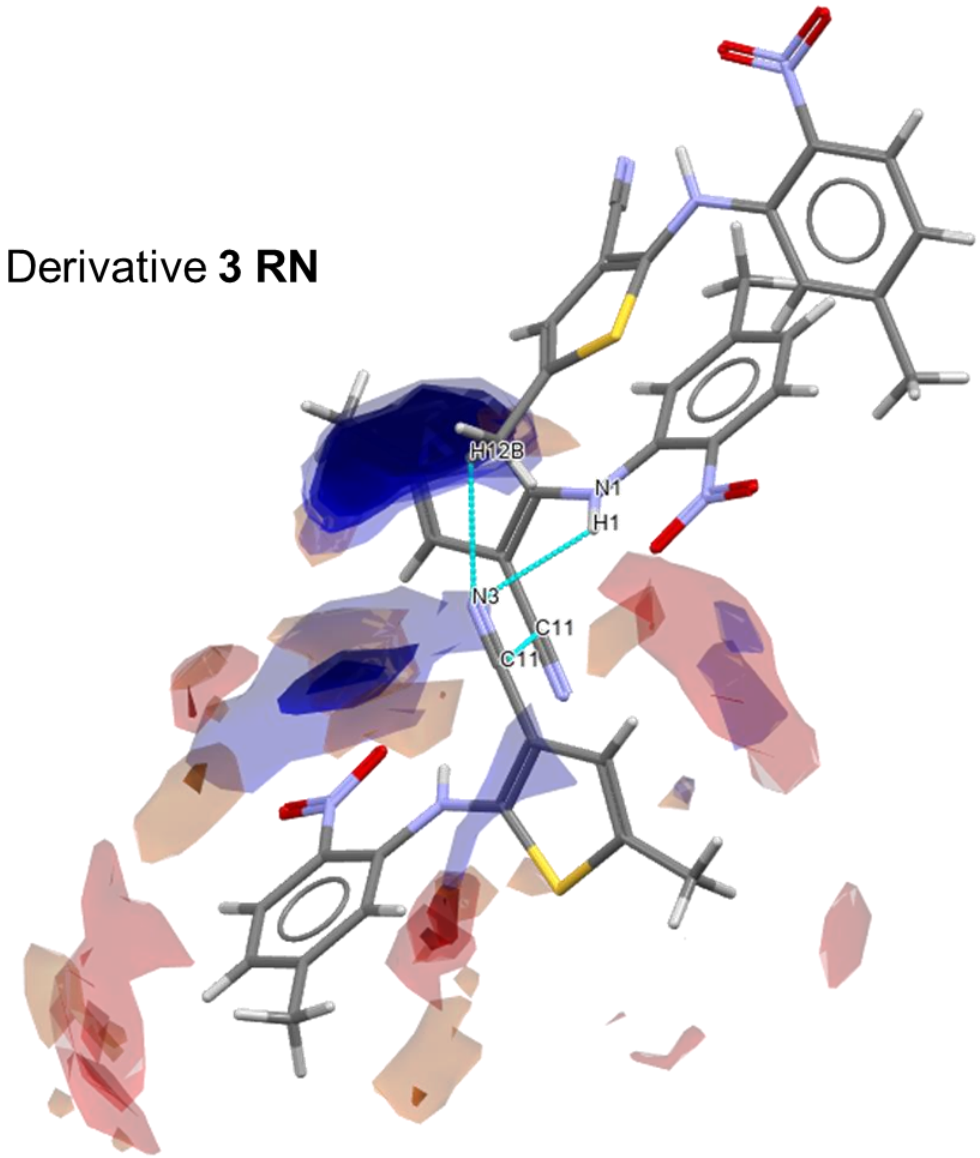


Figure 107; a visualisation of the intermolecular environment observed around the nitrile group in the crystal lattice of 3 RN.

Overall the intermolecular environment of **3 RN** is less than favourable, which suggesting the existence of more stable polymorphic forms of Derivative 3 which display more ideal intermolecular environments.

Red prisms of derivative **4** were isolated via slow evaporation from an excess of EtOAc. **4 RP** is a structural analogue of the ROY scaffold in which the C-H bond at the 6'-position of the nitrobenzene ring of ROY is replaced by at C-CH_3 bond. Figure 106 contains a visualisation of the intermolecular contacts formed within the crystal lattice of **4 RP**, along with FIMs calculated for **4 RP**. There a large number of contacts formed by molecules within the crystal lattice of **4 RP**, however a significant proportion of these contacts are located in/around their corresponding FIMs hotspots. There are also multiple π-π type interactions observed, similar to the previous structures discussed in this section.

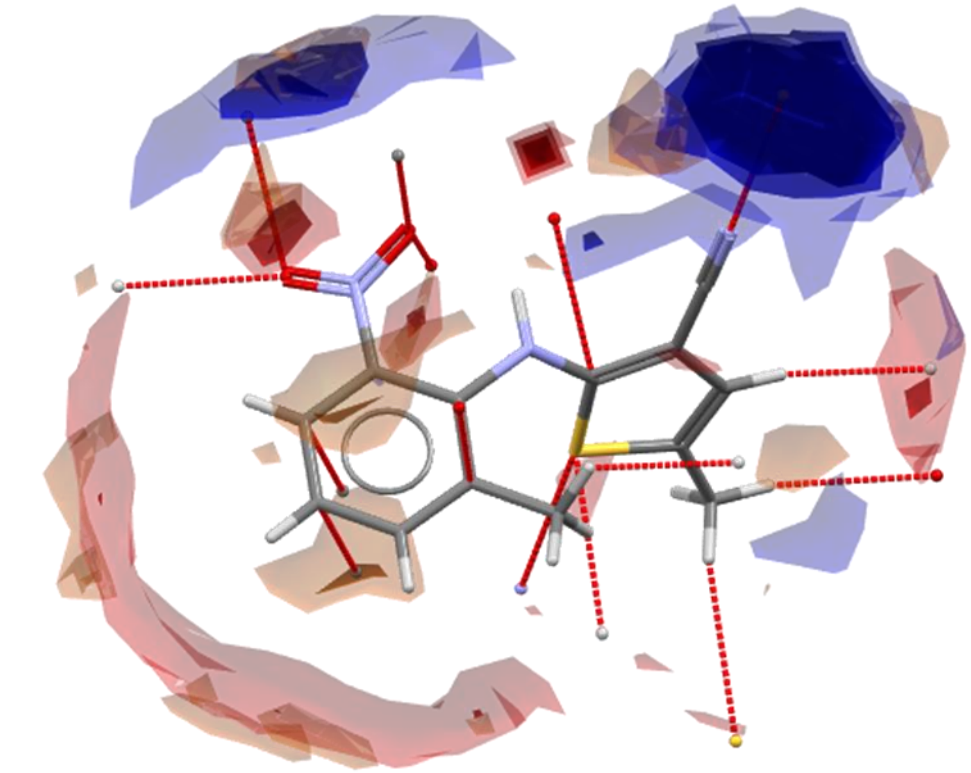


Figure 108; a visualisation of the overall intermolecular environment observed within the crystal structure of 4 RP.

The nitrile nitrogen of **4 RP** is involved in a single intermolecular contact located in the centre of the dark blue (uncharged NH hydrogen) hotspot associated with the CN group, which would suggest a favourable CN nitrogen environment. However, the partner for the CN nitrogen in this contact is the thiophene sulfur atom, which itself is involved in two separate intermolecular contacts, the other being with the thiophene-bound methyl H12B hydrogen. There are no significant FIMs hotspots associated with the thiophene sulfur atom, which could suggest interactions with the thiophene sulfur atom are unfavourable for **4 RP**. Nevertheless, the two contacts centred on the thiophene sulfur atom in **4 RP** suggest an unfavourable environment for the sulfur atom. Visualisations of the nitrile and sulfur environments of **4 RP**, along with FIMs calculated for **4 RP**, can be found in Figure 107 and Figure 108, respectively.

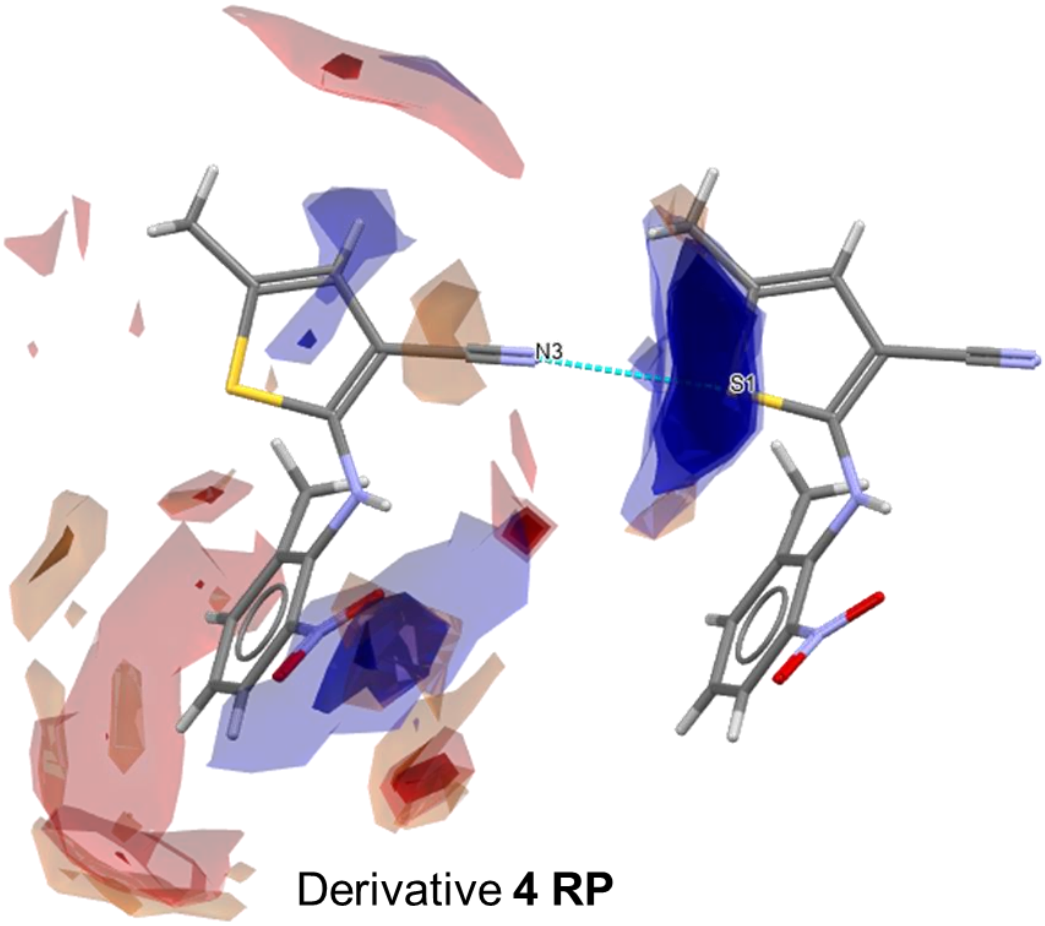


Figure 109; a visualisation of the intermolecular environment observed around the nitrile group within the crystal structure of 4 RP.

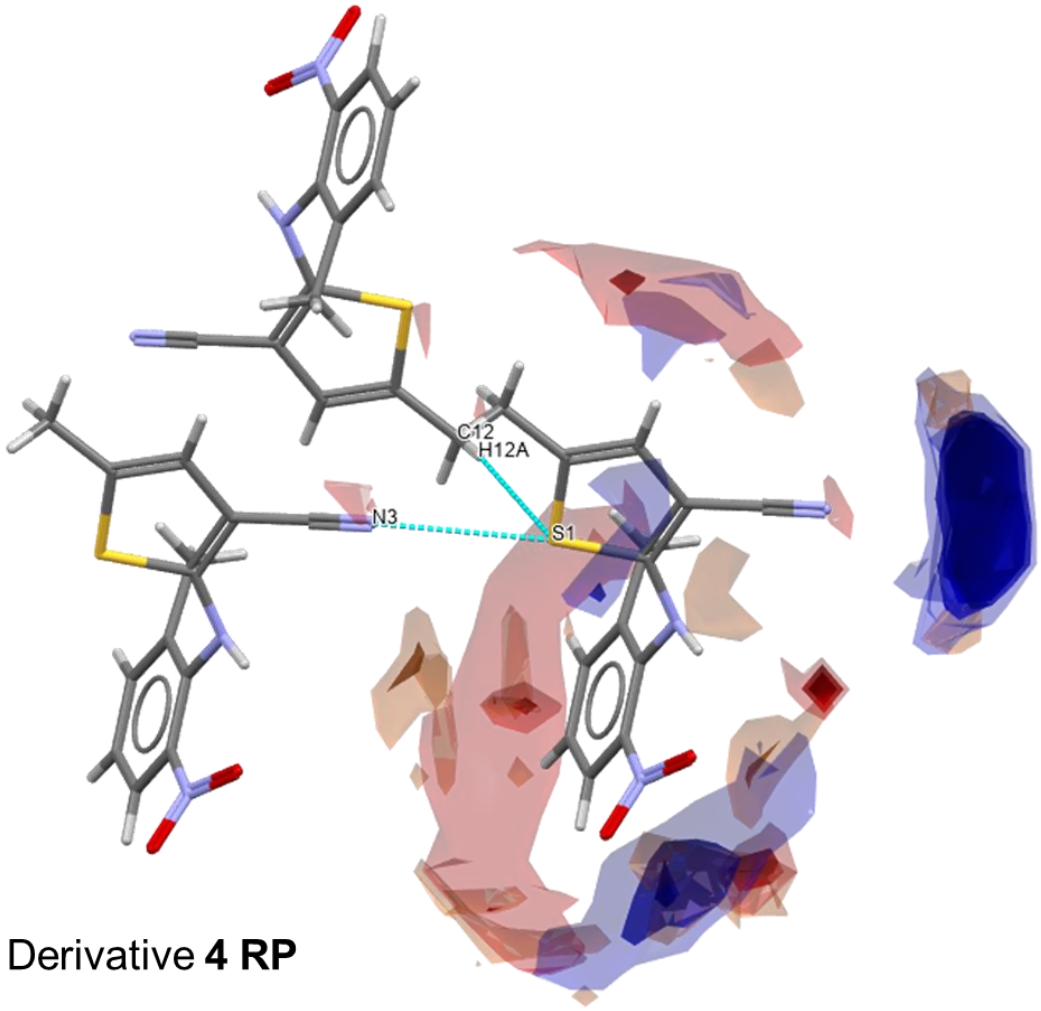


Figure 110; a visualisation of the intermolecular contacts formed between the thiophene S1 atom and both of the nitrile N3 and thiophene-bound methyl H12A hydrogen from within the crystal structure of 4 RP.

The two oxygen atoms, O1 and O2, of the nitro group in **4 RP** both form two intermolecular contacts, yet exist in very different environments. The O1 atom is involved in two contacts above and below the plane of the nitro group, with the nitrobenzene C10 and nitro O1 atoms of neighbouring **4 RP** molecules. The O2 atom is involved in a contact with the thiophene C1 carbon which is in the centre of the dark blue (uncharged NH hydrogen) hotspot, along with another contact with the thiophene-bound methyl H12B hydrogen which is separate from any of the hotspots associated with the nitro group of **4 RP**. This makes evaluating the favourability of the intermolecular environment of the nitro group difficult. A visualisation of the intermolecular environment of the nitro group in **4 RP** can be found in Figure 109.

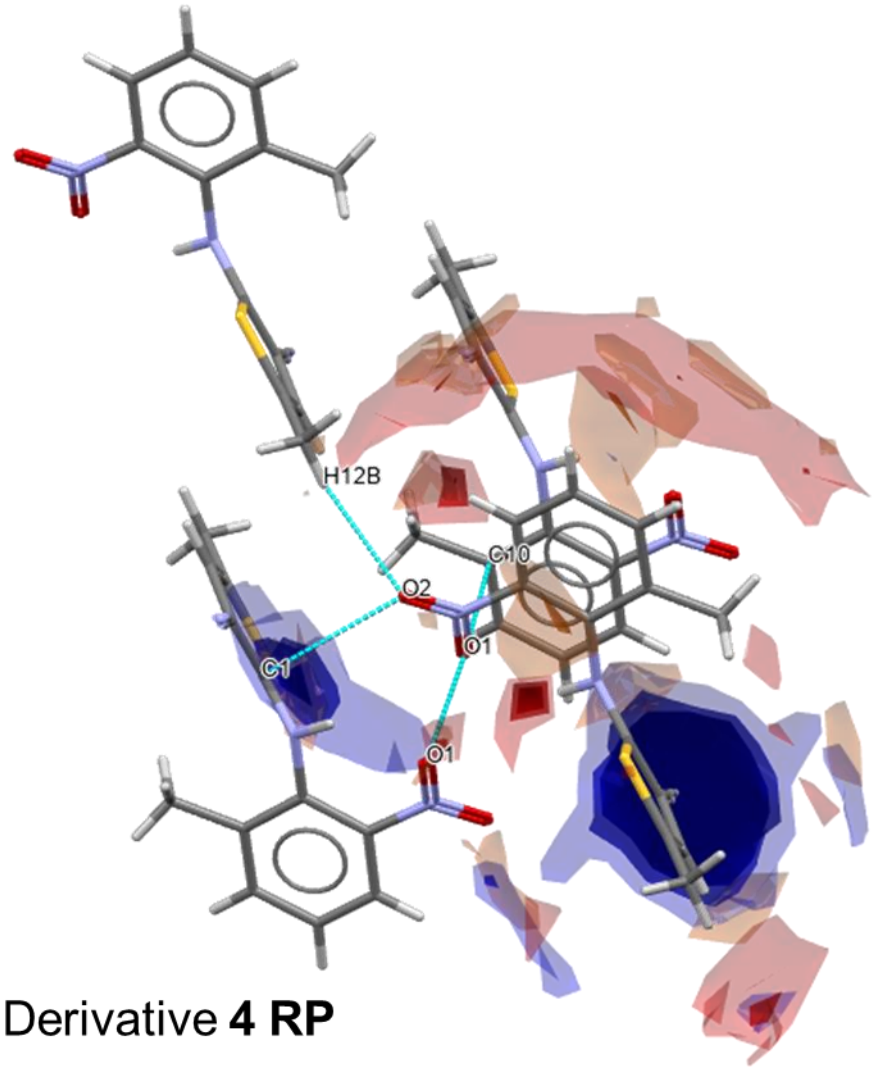


Figure 111; a visualisation of the nitro environment observed within the crystal structure of 4 RP.

Overall there are a number of conflicting features of the intermolecular environment present in **4 RP**. The nitrile N3 atom is involved in a single contact located in the centre of the dark blue (uncharged NH hydrogen) hotspot associated with the nitrile group, which, despite the unusual S1 partner, suggests a favourable nitrile environment. The nitro group presents a more difficult environment to evaluate, but the location of the contacts between the two neighbouring O1 atoms, and the O2 and C1 atoms, suggest a degree of favourability of the nitro environment. The relative influence of the nitrile group on the stability of ROY forms in Chapter 2.2 suggests that **4 RP** would be a relatively stable form. However, the overall intermolecular environment presented by **4 RP** suggests that there could be more stable forms accessible to Derivative **4**.

3.5.2 Derivatives 5 – 8 Intermolecular Environment

Derivatives **5** – **8** were designed using the same design principles as Derivatives **1** – **4**, using the 2015 study of the fenamate polymorphophore by Lopes-Meijas et al. as inspiration, however instead of replacing the nitrobenzene C-H bonds with C-CCH_3 bonds, they are replaced with C-Cl bonds instead. The methyl and chloro groups are similar in volume (19 A^3 vs 24 A^3), yet display different electronic properties, hence it is theorised that the chloro groups could influence the intermolecular environment of ROY derivatives through different ways to the methyl groups discussed in 3.3.1.

Red needles of derivative **5** were isolated via slow evaporation from excess EtOAc. **5 ON** is a derivative of the ROY scaffold in which the C-H bond at the 3' position of the nitrobenzene ring is replaced by a C-Cl bond. A summary of the overall intermolecular observed within the crystal structure of **5 ON** can be found in Figure 110. The overall environment presented by **5 ON** looks relatively favourable; there is a low number of intermolecular contacts observed, and over half of the observed contacts are located at the edge of, or within their corresponding FIMs hotspots, which is reminiscent of the stable ROY polymorphs **ON** (QAXMEH) and **Y** (QAXMEH01) (see Figure 46 and Figure 47, respectively.). However, analysis of the local environments of the nitrile and nitro functional groups reveals a number of less favourable features.

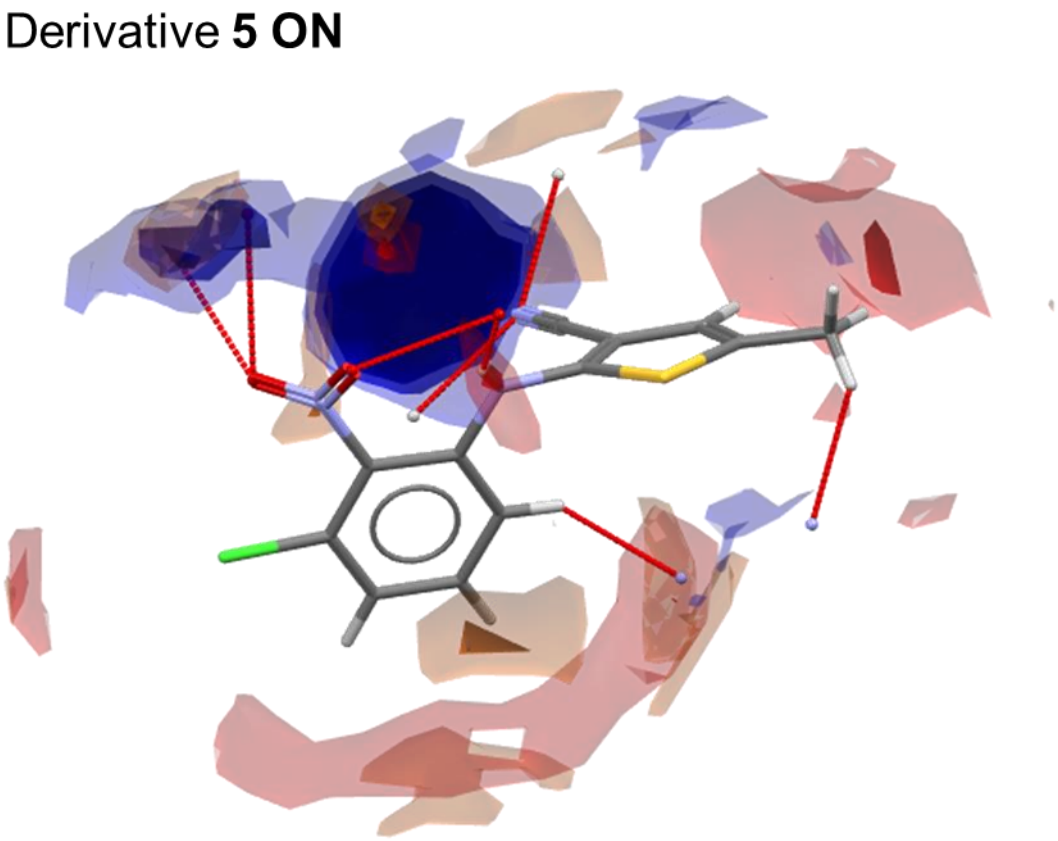


Figure 112; a visualisation of the overall intermolecular environment observed within the crystal structure of 5 ON.

The intermolecular environment of the nitrile group in **5 ON** clearly displays some unfavourable features, highlighted by the two intermolecular contacts involving the nitrile N3 atom which are located outside of the nitrile FIMs hotspots. The N3 atom forms one contact with the nitrobenzene-bound H10 hydrogen located at the 6' position of the thiophene ring, and another with the thiophene-bound methyl H12C hydrogen, with both of these contacts being observed at the edge of the blue (uncharged NH hydrogen) hotspot associated with the nitrile group. The contact between the nitrile N3 and aromatic H10 hydrogen could be considered favourable, but the presence of the additional contact between N3 and the methyl H12C hydrogen suggests and unfavourable nitrile environment overall. A visualisation of the nitrile environment for **5 ON** can be found in Figure 111.

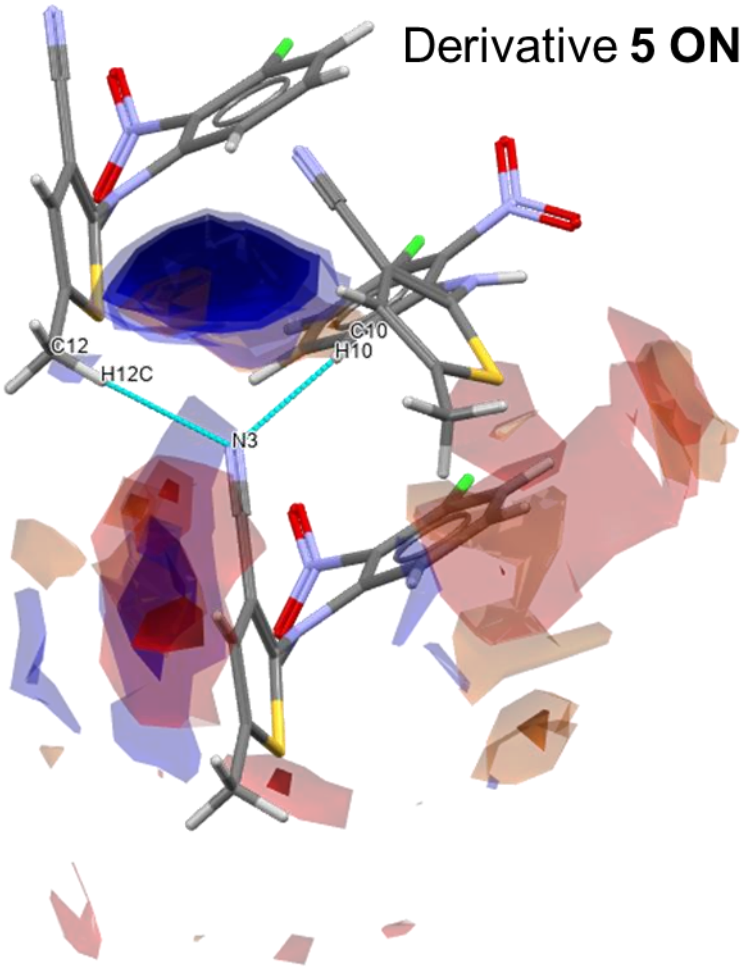


Figure 113; a visualisation of the nitrile environment observed within the crystal structure of 5 ON.

The nitro environment in **5 ON** contrasts with the nitrile environment due to the location of the observed contacts relative to the corresponding FIMs hotspots. The O2 atom of the nitro group is involved in a contact with the amino H1 hydrogen, which is located in the blue (uncharged NH hydrogen) and red

(carbonyl oxygen) hotspots associated with the O2 and N1 groups, respectively, and could be considered a favourable interaction. However, due to the inherent structure of the nitro group, this interaction also involves the O1 atom, which is involved in a contact with the O2 atom, reducing the overall favourability of the nitro environment. A visualisation of the nitro environment in **5 ON** can be found in Figure 112.

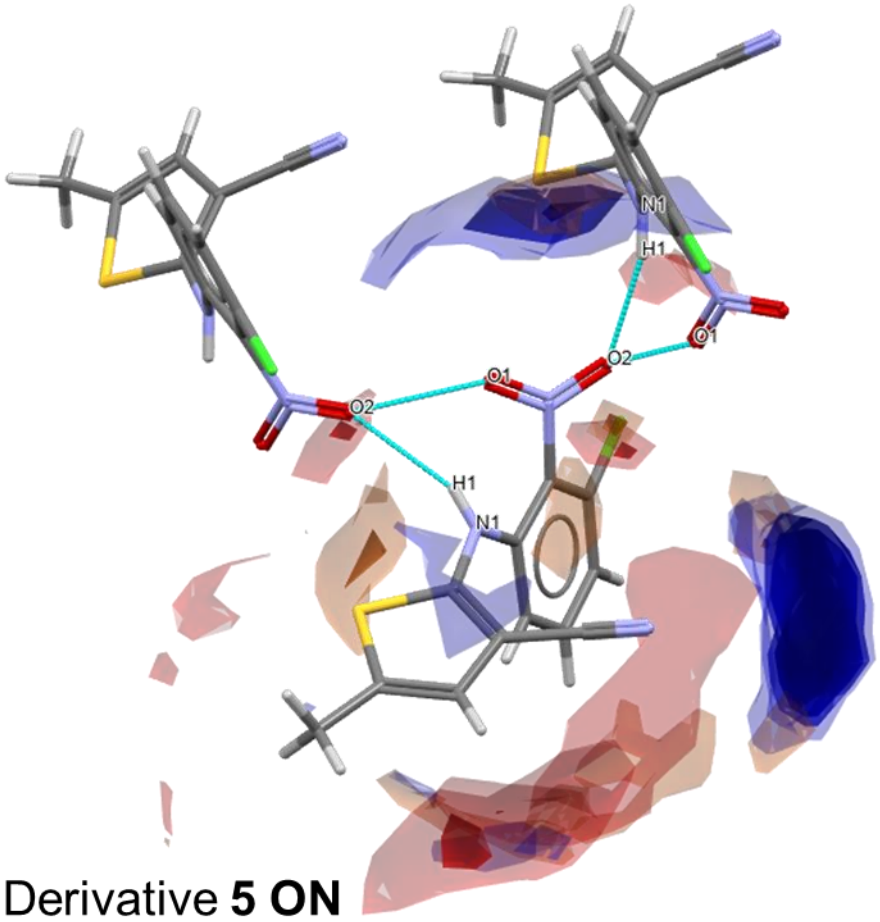


Figure 114; a visualisation of the nitro environment observed within the crystal structure of 5 ON.

Despite the overall environment appearing similar to the stable ROY Forms **Y** (QAXMEH01) and **O** (QAXMEH03), the nitrile environment of **5 ON** suggests a less stable form overall. The presence of the O2 – H1-N1 H-bond was indicative of unfavourable intermolecular environments, however the location of the O2 – H1-N1 contacts in **5 ON** are more favourable than those observed in the **YN** (QAXMEH04) Form, which could suggest a more stable environment for the **5 ON** Form. Nevertheless, the overall environment observed for **5 ON** suggests that more stable polymorphs of **5** are likely.

Red needles of derivative **6** were isolated via slow evaporation from an excess of EtOAc. It was shown in section 3.2.2 that the crystal structure for the red needles

of derivative **6 RN** was poorly resolved, with an R^2 value (describing the difference between the observed structure and an ideal model of the predicted crystal lattice) of 8.3%, significantly above the 5% minimum threshold for the value. This suggests that any conclusions inferred from SFI analysis of the **6 RN** crystal structure are unreliable. This fact is reinforced when visualising the overall FIMs environment of **6 RN** (Figure 113). The large number of contacts centred on the nitro, chloro and methyl groups result in an intermolecular environment unlike any of the ROY polymorphs or derivatives discussed previously. As such a full analysis of the intermolecular environment observed in **6 RN** is moot, and is included here for completeness.

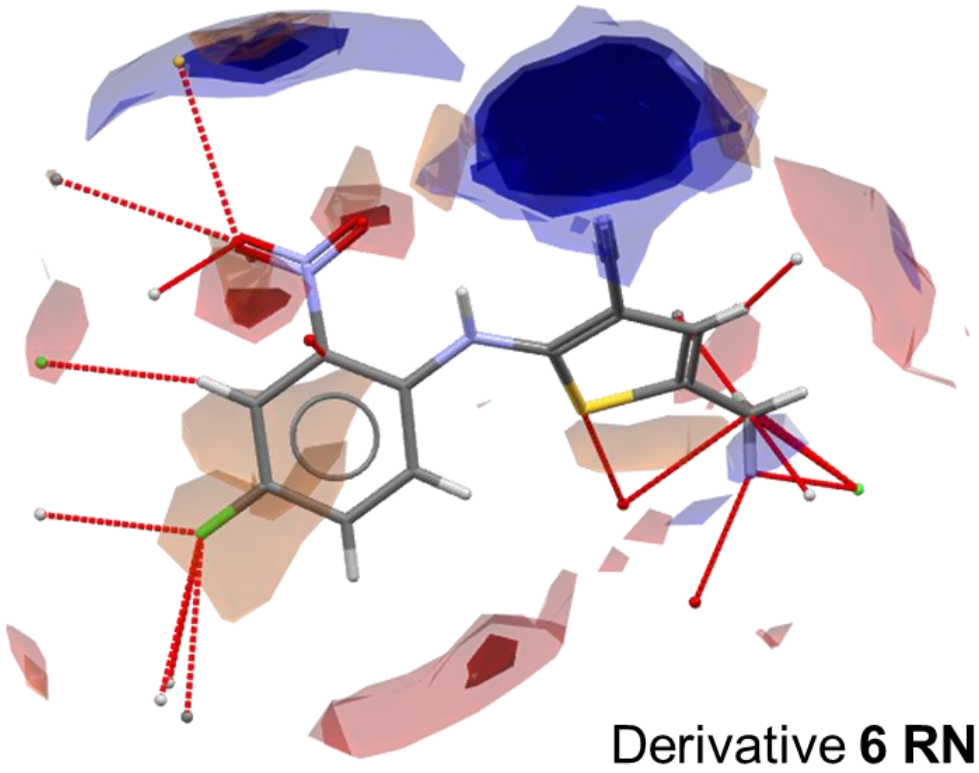


Figure 115; a visualisation of the overall intermolecular environment observed within the crystal structure of 6 RN.

Despite the low resolution of the **6 RN** form resulting in a crystal structure for which any trends observed must be taken with a pinch of result/could be considered unreliable, there are noticeable similarities between the intermolecular environment of **6 RN** and the yellow prisms of Derivative **7**.

Yellow prisms of derivative **7** were isolated via slow evaporation from excess EtOAc. The resolution of the crystal structure of **7 YP** is below the required threshold, at 3.07 %, and as such the unusual environment of observed for **7 YP** cannot be dismissed as a result of a low-quality crystal structure. A visualisation

of the overall intermolecular environment observed in **7 YP** can be found in Figure 114.

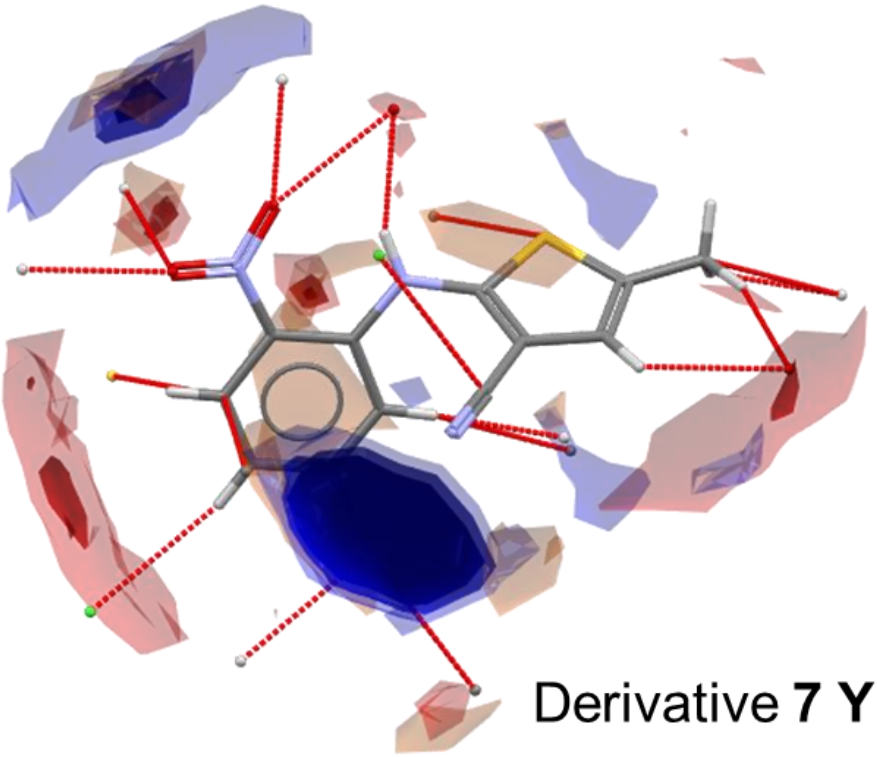


Figure 116; a visualisation of the overall intermolecular environment observed within the crystal structure of 7 Y

The first noteworthy feature of the intermolecular environment of **7 YP** is the lack of any contacts involving the nitrile N3 nitrogen, resulting in the dark blue (uncharged NH hydrogen) hotspot associated with the nitrile group devoid of any contacts. The next feature of note is the environment of the nitro O1 and O2 atoms, for which a visualisation can be found in Figure 115. Both O1 and O2 atoms in **7 YP** are involved in two contacts each. The nitro O1 atom is involved a contact with the amine H1 proton, similar to the environment of the nitro group in **5 ON**. However, the O2 atom is involved in contacts with the thiophene-bound H3 and methyl H12B hydrogens. The location of the N1-H1 – N3 H-bond relative to the corresponding Fims hotspots of the amine and nitro groups suggests a favourable interaction, but the O2 environment coupled with the bifurcation of both nitro oxygen atoms suggests an unfavourable nitro environment overall for **7 YP**.

The intermolecular environment observed in **7 YP** has more in common with the metastable ROY Forms discussed in 2.2.1.2 than with the stable Forms discussed in 2.2.1.1, thus it is predicted that **7 YP** is a metastable Form of **7**, and more stable forms of **7** would be identified via polymorph screen.

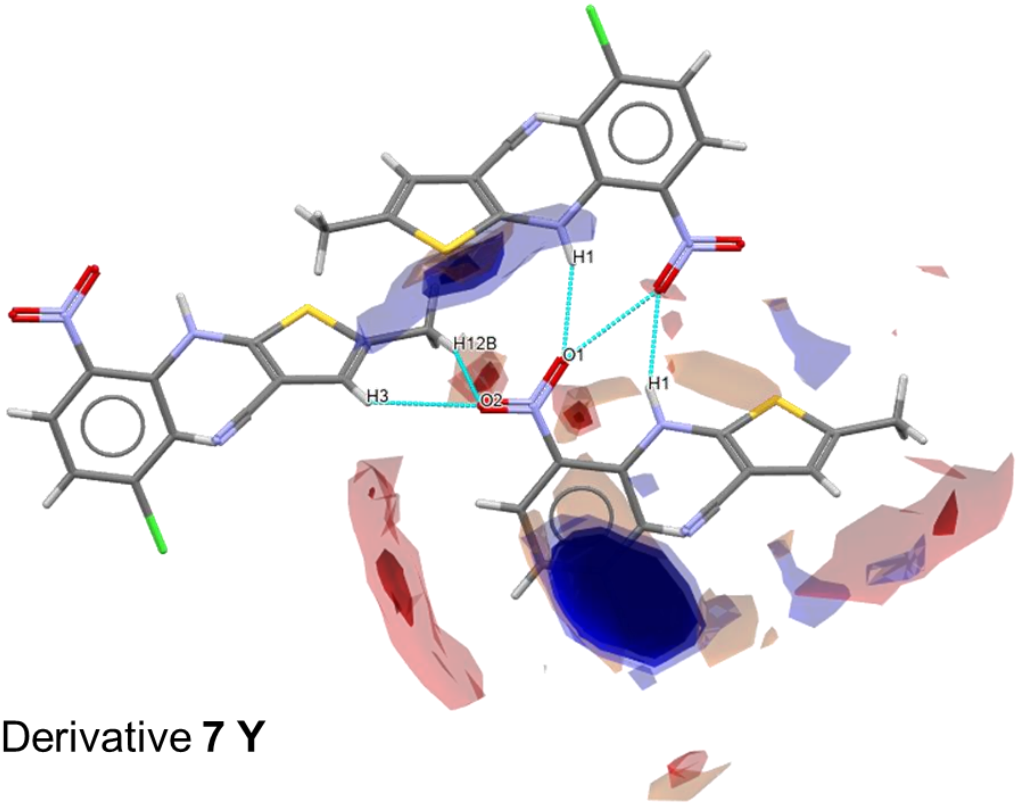


Figure 117; a visualisation of the nitro environment observed within the crystal structure of 7 Y.

The final derivative to discuss in this section derivative **8**, which was crystallised from an excess of EtOAc to yield red needles (**8 RN**). A visualisation of the overall intermolecular environment of **8 RN** can be found in Figure 116. Similar to the intermolecular environments observed for **6 RN** and **7 YP**, it is clear at glance that the intermolecular environment in 8 RN is less than ideal when compared to the stable ROY forms discussed in 2.2.1.1. The lack of contacts formed by the nitrile group and one half of the nitro group are the most noteworthy issues. However the lack of any intermolecular contacts located in/around FIMs hotspots throughout the entirety of the intermolecular environment of **8 RN** is another cause for concern.

Overall, it is clear through comparison of the intermolecular environment observed in **8 RN** to the stable and metastable ROY Forms discussed in section 2.2 that the intermolecular environment of **8 RN** is less than ideal. As such, it is highly likely that **8 RN** is a metastable form of Derivative **8**, and that a sufficient polymorph screen would identify more stable forms of **8**.

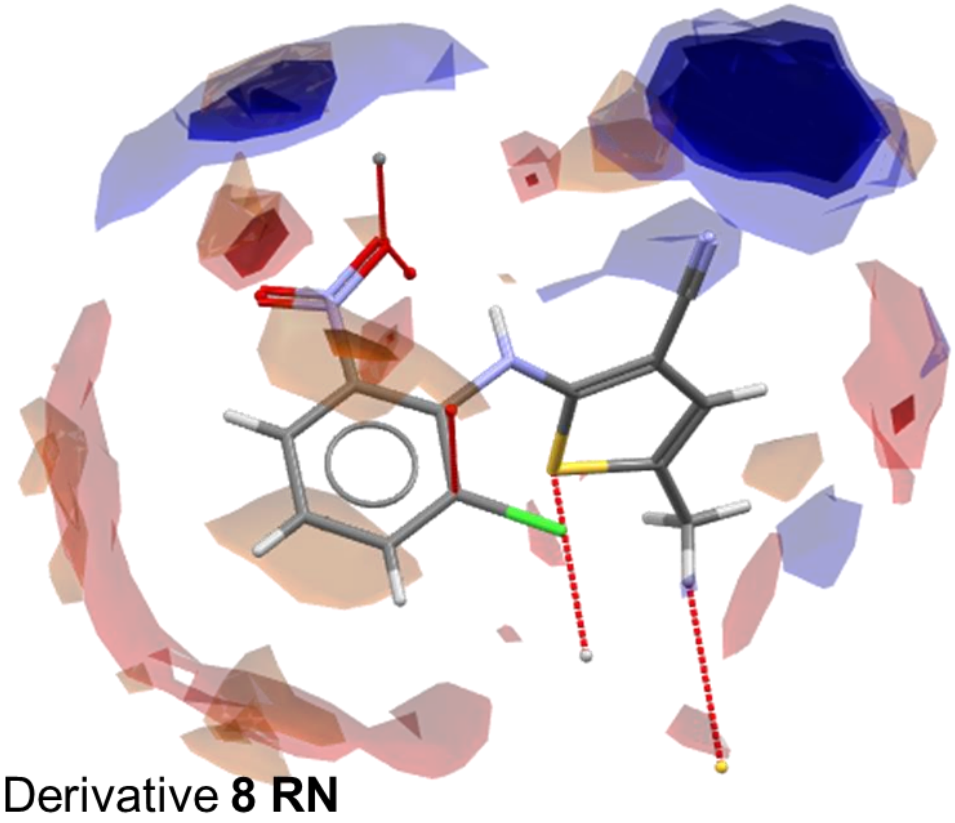


Figure 118; a visualisation of the overall intermolecular environment observed within the crystal structure of 8 RN.

3.5.3 Derivatives 9 and 10 Intermolecular Environment

Derivatives **9** and **10** were designed following the same design philosophy as Derivatives **1** – **8**, however, the additional complexities relating to the synthesis of **9** and **10**, along with their unique crystallographic properties, necessitated a separate section for the discussion of **9** and **10**. Both **9 DR** and **10 RP** required a significant number of experiments in order to generate crystals of sufficient quality for SC-XRD, which is unusual for ROY and the ROY derivatives generated for this work. Moreover, it was discovered in 3.2.3 that both **9 DR** and **10 RP** display intramolecular torsions which are uncommon for ROY-like molecules within the CSD and synthesised for this work, with the qthio torsion values for **9 DR** and **10 RP** being 0.34 and 0.52, respectively, significantly more planar than ROY-like molecules discussed previously.

Dark red needles of Derivative **9** were generated via slow evaporation of a solution of **9** in EtOAc. A visualisation of the overall intermolecular environment observed within the crystal structure of **9 DR** can be found in Figure 117. At a glance the intermolecular environment within **9 DR** looks reasonably favourable, with contacts involving the nitrile and nitro groups being located in or at the

periphery of their corresponding FIMs hotspots. The relatively low number of remaining contacts are located close to their corresponding FIMs hotspots, suggesting a relatively favourable intermolecular environment overall.

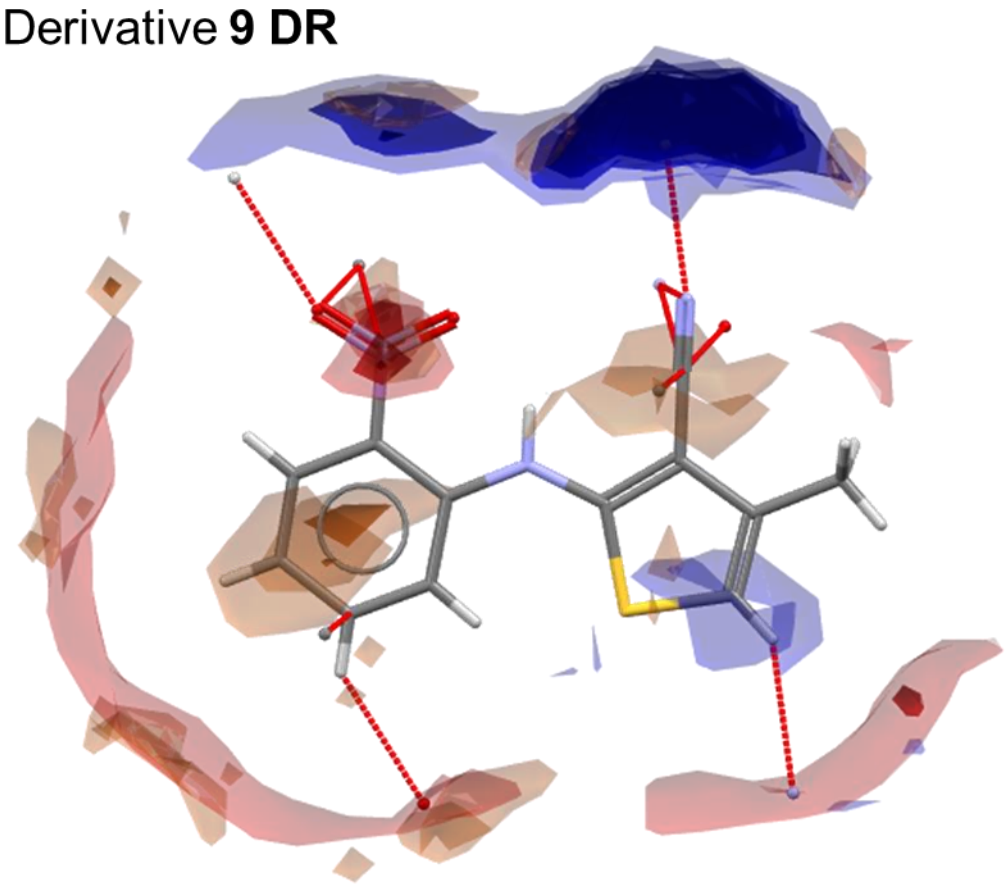


Figure 119; a visualisation of the overall intermolecular environment observed within the crystal lattice of 9 DR.

However, despite the presence of favourable contacts involving the nitrile and nitro groups, there are a number of features which may have an adverse effect on the intermolecular environment. For the nitrile environment, the nitrile N3 atom is involved in two contacts, the first is a favourable contact with the thiophene H4 hydrogen located in the centre of the dark blue (uncharged NH hydrogen) hotspot associated with the N3 atom. The second contact is below the plane of the nitrile group, and is formed with the nitro N2 and O2 atoms. This contacts looks like a π-type interaction, and is located in the brown (aromatic CH carbon) hotspot associated with the nitrile group, which could suggest a favourable interaction. However, the presence of two contacts involving the nitrile N3 atom is less than ideal. A visualisation of the nitrile environment can be found in Figure 118. This suggests a favourable intermolecular environment for the nitrile group in **9 DR**, though not the ideal environment.

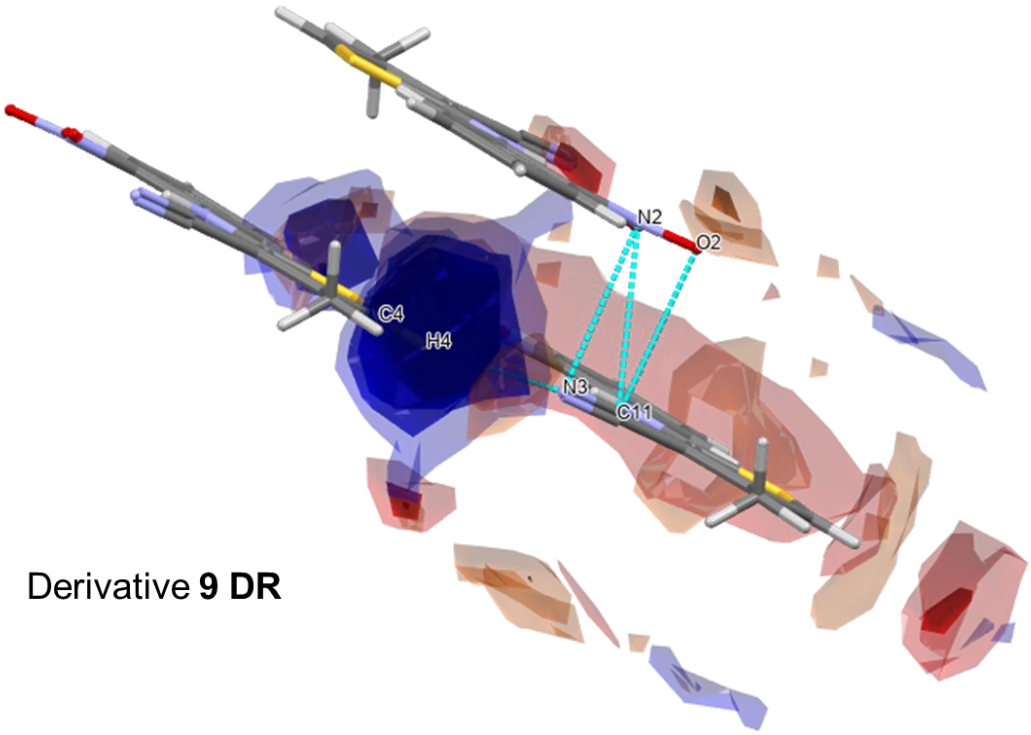


Figure 120; a visualisation of the nitrile environment observed within 9 DR.

The nitro environment is similar to that observed for the nitrile group in **9 DR**, partially due to the two groups being involved in the π-type interaction discussed above. The nitro O2 atom is also involved in a single contact with the nitrobenzene H9 hydrogen, located at the edge of the blue (uncharged NH hydrogen) hotspot associated with the nitro group, suggesting a favourable interaction. Moreover, the π-type interaction with the nitrile group is located within a mixed brown (aromatic CH carbon) and dark red (carbonyl oxygen) hotspot located above the plane of the nitro group, suggesting this interaction is favourable for the nitro group. A visualisation of the intermolecular environment around the nitro group in **9 DR** can be found in Figure 119.

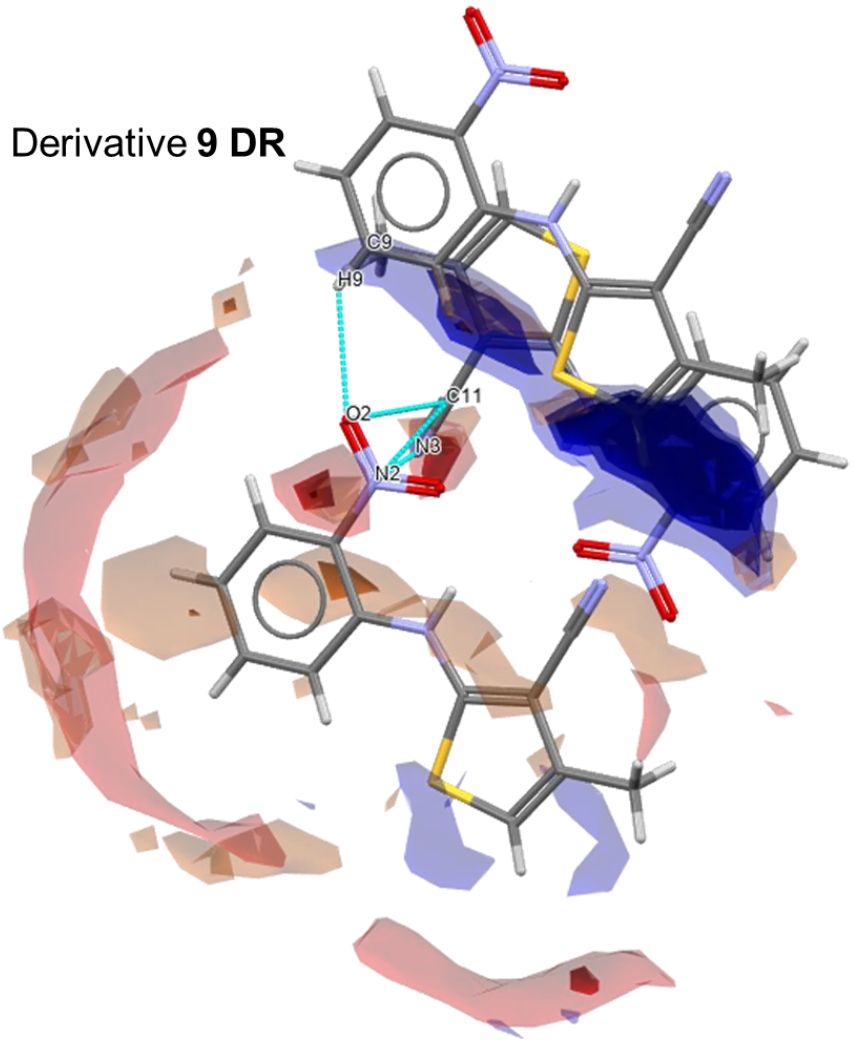


Figure 121; a visualisation of the nitro environment observed within 9 DR.

Overall, the intermolecular environment observed within the crystal structure of **9 DR** is relatively favourable, containing a small number of contacts located within, or close to, their corresponding FIMs hotspots. This suggests that **9 DR** is a stable polymorph of **9**, however, it could be possible that a more stable structure could be found through polymorph screening.

Red prisms of Derivative **10** were isolated via the slow evaporation of a solution of derivative **10** in EtOAc. A visualisation of the overall intermolecular environment present within the crystal structure of **10 RP** can be found in Figure 120. The overall intermolecular environment for **10 RP** looks very favourable, with nitro and nitrile environments similar to those observed for the **ON** (QAXMEH) polymorph of ROY, one of the most stable ROY forms (reference the overall **ON** environment from 2.2). It should also be noted that, similar to **9 DR**, there seems to be a significant number of π-type interactions, highlighted by the large number of contacts above and below the plane of the aromatic rings in **10 RP**. This is likely due to the planarity of the **10** molecule within **10 RP** favouring the formation of π-π stacking interaction.

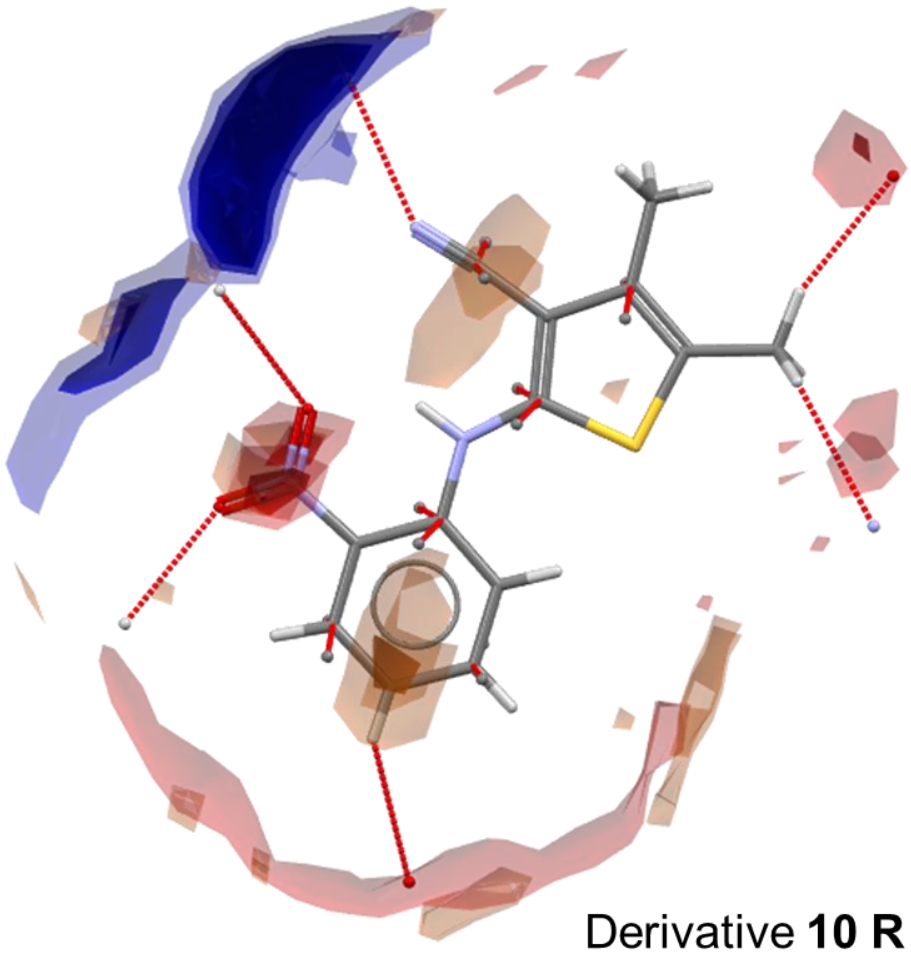


Figure 122; a visualisation of the overall intermolecular environment observed within the crystal lattice of 10 RP.

Alongside the π-type interactions, the nitrile N3 atom in **10 RP** forms a contact with the thiophene-bound methyl H12A hydrogen. Whilst the aliphatic H12A proton isn't the ideal partner for the N3 atom, the location of the contact within the prominent dark blue (uncharged NH hydrogen) hotspot associated with the nitrile group suggests this is a favourable contact. Furthermore, this thiophene-bound methyl group is involved in another contact with the nitro O2 atom. Similarly, the H12B proton isn't the ideal partner for the nitro O2 atom, but the location of the contact at the periphery of the light blue (uncharged NH hydrogen) hotspot associated with the O2 atom suggests this contact is favourable.

3.5.4 Derivative 11 Intermolecular Environment

Derivative **11** along with Derivative **2** are unique amongst the analogues synthesised for this section as they are the two structures which had been previously reported within the literature. **11** was originally discovered and characterised by Leva-Perez et al in 2010 as part of a review of the industrial routes towards the production of olanzapine. Colourless blocks of **11** were

characterised via SC-XRD and uploaded to the CSD under the refcode XABXEF. After the synthesis and characterisation of **11** performed for this work yielded similar colourless blocks, cross comparison of the structural information produced for this work and that of XAABXEF revealed that **11 CB** and XABXEF are the same form of **11**. The fact that only one form of **11** is found within the CSD, and that this work yielded the exact same crystal form of **11**, suggests that **11 CB**/XABXEF could be the most stable form of **11**. This could also suggest that **11** is monomorphic, or at least significantly less prone to polymorphism that other ROY-like structures. Analysis of the intermolecular environments observed within the crystal structures of **11 CB**/XABXEF could provide a rationalisation for the perceived stability of the **11 CB** form.

A visualisation of the overall FIMs environment observed within the lattice of **11 CB** can be found in Figure 121. Replacement of the NO_2 group with the NH_2 group in 11 has a significant effect on the FIMs hotspots generated for 11 CB. The broad and diffuse blue (uncharged NH hydrogen) hotspots associated with the NO_2 group are replaced with very dark red (carbonyl oxygen) hotspots in line with the protons of the NH_2 group, along with a dark blue (uncharged NH hydrogen) hotspot located above the plane of the nitrogen atom of the NH_2 group in **11 CB**. The removal of the NO_2 group in **11 CB** also has the effect of removing the intramolecular H-bond involving the bridging NH group, which results in a significantly darker red (carbonyl oxygen) hotspot associated with the birding amine in **11 CB**.

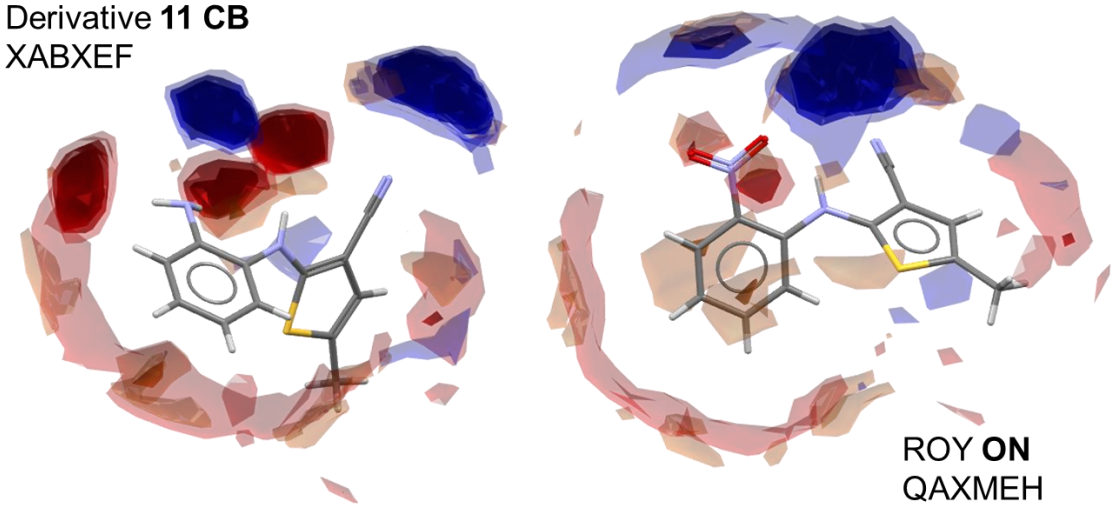


Figure 123; a comparison of the FIMs hotspots calculated for 11 CB and the ON (QAXMEH) form of ROY

By visualising the short contacts in conjunction with the FIMs hotspots for **11 CB** the origin of the stability of the **11 CB** form becomes more apparent. A visualisation of the overall intermolecular environment within the crystal lattice of

11 CB can be found in Figure 122. The first thing to notice for **11 CB** is the significant roles played by all three atoms of the NH_2 group introduced during the synthesis of **11**, along with the bridging amine proton. Contacts involving the primary amine N2 atom, both H2A and H2B protons of the primary amine, and the H1 proton of the bridging amine are all involved in contacts located inside their corresponding FIMs hotspots, suggesting favourable environments for all atoms mentioned. The only real cause for concern within the crystal lattice of **11 CB** is the nitrile N3 atom, which is involved in 3 contacts, which could suggest a less favourable nitrile environment.

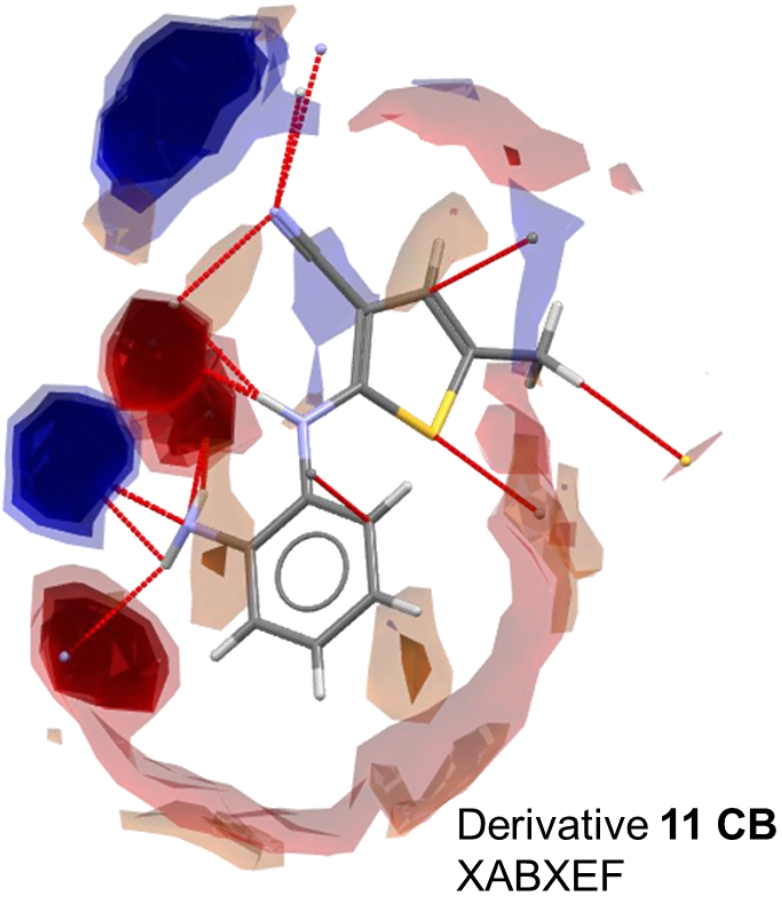


Figure 124; a visualisation of the overall intermolecular environment observed within the crystal lattice of 11 CB.

Another interesting feature of the intermolecular environment of **11 CB** becomes apparent when the contacts involving the nitrile and primary amine groups are expanded. A number of H-bonds connecting the nitrile N3 atom to both of the primary amine H2A and H2B protons is present, located within the dark red (carbonyl oxygen) hotspots associated with the primary amine and toward the edge of the blue (uncharged NH hydrogen) hotspot associated with the nitrile group. Alongside a H-bond present between the N2 nitrogen of the primary amine

with the H1 proton of the bridging secondary amine, these H-bonds form somewhat of a chain throughout the **11 CB** structure. This could add long range order within the crystal lattice of **11 CB**, therefore further increasing the relative polymorphic stability of the **11 CB** form. A visualisation of the H-bonds present within the **11 CB** crystal lattice can be found in Figure 123.

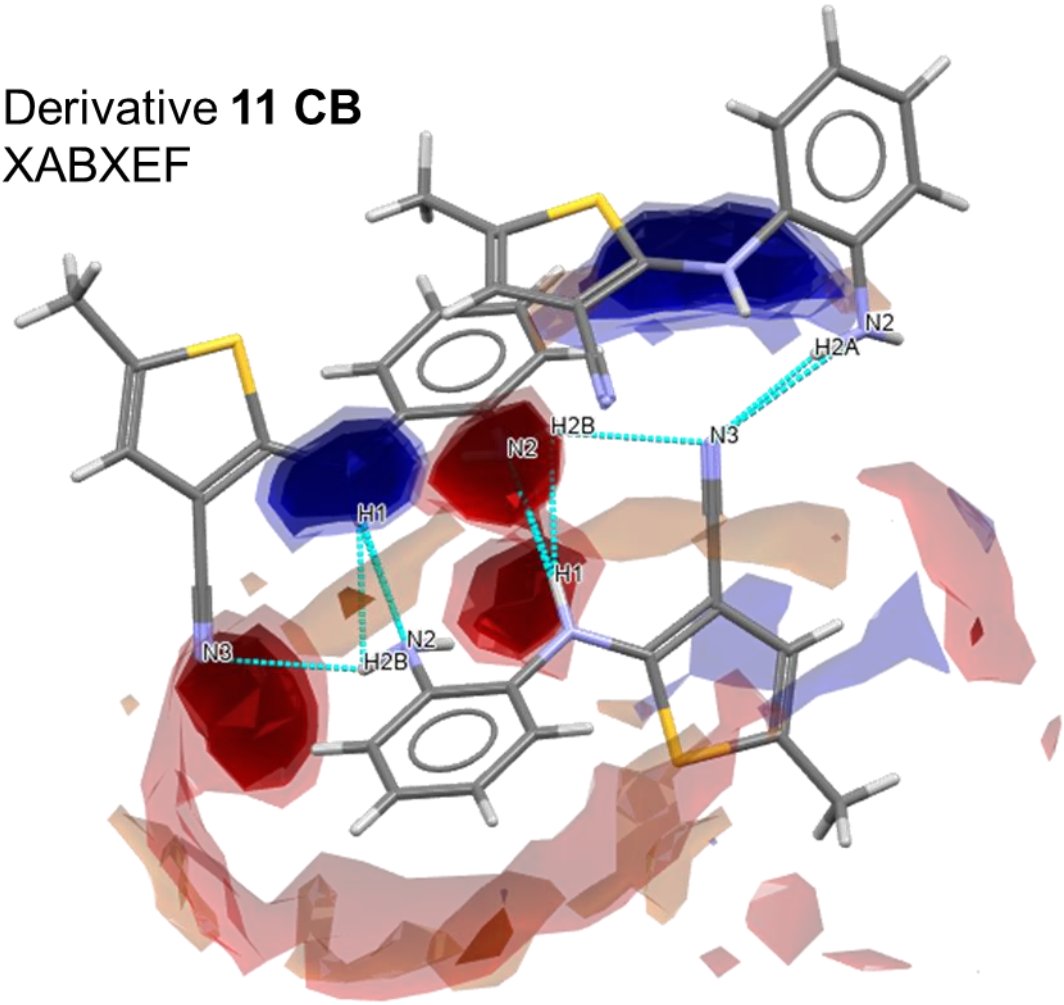


Figure 125; a visualisation of the H-bonds observed within the crystal lattice of 11 CB.

The overall picture of the intermolecular environment within **11 CB** helps rationalise the perceived stability and potential monomorphic nature of Derivative **11.** The FIMs produced by the chemical structure of **11** within **11 CB** result in significant differences to the ROY polymorphs and ROY derivatives previously discussed in this work, which may explain why a favourable NH_2 environment is formed preferably to a favourable CN environment in **11 CB**, whilst maintain good polymorphic stability. It would be interesting to see if any other forms of **11** are discovered via polymorph screen, and how the intermolecular environment of these forms may differ to **11 CB**.

3.5.5 Derivative 12 ON and 12 RN Intermolecular Environment

There were initially some difficulties in generating high quality single crystals of **12**, and a number of experiments utilising a range of solvents were employed in order to access crystals of sufficient quality. During these experiments, it was noted that crystals grown from an excess of EtOAc yielded thin red needles, whilst similar experiments performed in chloroform yielded orange needles. Structural determination via SC-XRD revealed that these two samples are distinct crystal forms, referred to as **12 RN** (red needles) and **12 ON** (orange needles) herein. These two distinct forms of **12** provide us with an opportunity to utilise the SFI workflow, particularly analysis of the intermolecular environment, to evaluate which of these two forms is likely to be the most thermodynamically stable, which could then be verified via thermal analysis.

Derivative **12** is a derivative of the ROY scaffold in which the nitrile group at the 2 position of the thiophene ring is reduced to the corresponding aldehyde. Chapter 2.2 revealed that the nitrile group plays a key role in the stabilisation of the crystal lattices of the known ROY forms from within the CSD, with a favourable nitrile environment being a strong indicator of an overall favourable intermolecular environment. The introduction of the aldehyde group introduces a second intramolecular H-bond with the bridging amine proton within the structure of **12**, which can be seen in Figure 124. This intramolecular H-bond may have the effect of reducing the potential for the formation of intermolecular contacts with the aldehyde group due to the sharing of electrons within the intramolecular H-bond.

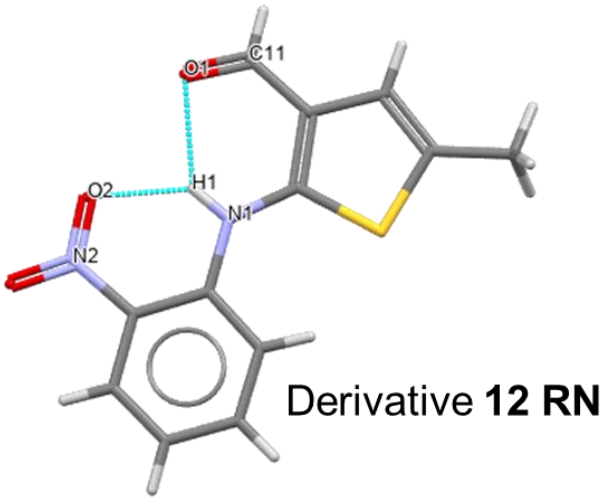


Figure 126; the bifurcated bridging amine group in 12 ON, featuring intramolecular H-bond with the nitro O2 and aldehyde O1 atoms

The replacement of the nitrile group with the aldehyde group in **12** leads to a smaller change in the FIMs environment of **12** than was observed for **11 CB** in 3.3.4, with the primary difference being the introduction of a dark red (carbonyl oxygen) hotspot associated with the aldehyde H11 proton. The dark blue

(uncharged NH hydrogen) hotspot associated with the aldehyde O1 atom is also smaller than the same hotspot for the nitrile nitrogen observed for other ROY-like structures discussed in Chapters 2 and 3. This reduction in FIMs hotspot area could signify a lesser role for the aldehyde oxygen in the stabilisation of **12** relative to the nitrile nitrogen in other ROY-like structures. Moreover, the smaller blue (uncharged NH hydrogen) hotspot associated with the aldehyde oxygen in **12** increases the relative size of the blue (uncharged NH hydrogen) hotspot associated with the nitro group in **12**, which could suggest a bigger role for the nitro group in the stabilisation of **12** forms. A comparison of the overall FIMs environments for **12 RN** and the **ON** (QAXMEH) form of ROY can be found in Figure 125.

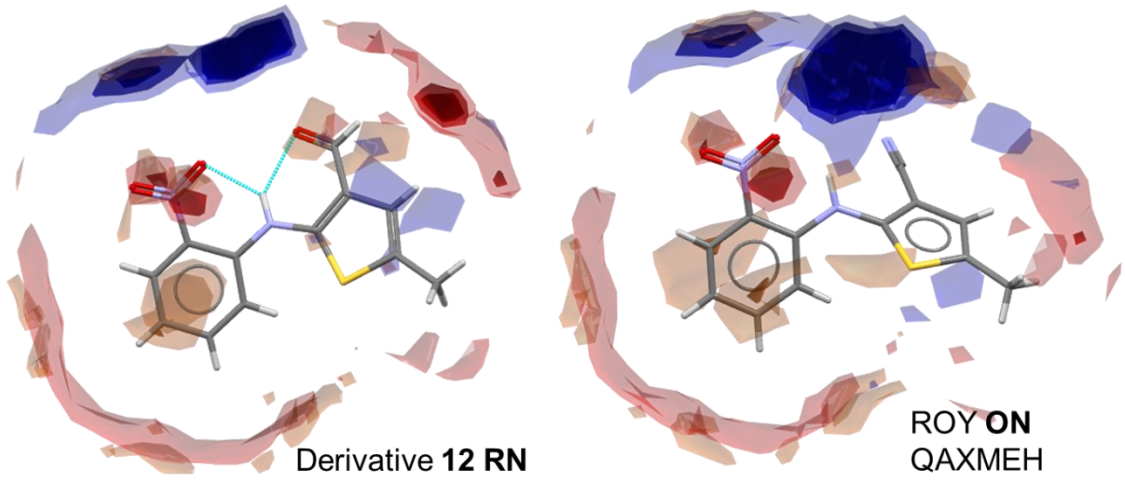


Figure 127; comparison of the FIMs calculated for 12 RN and the ON (QAXMEH) form of ROY.

The first form of **12** discovered was **12 ON**, which was found to have a melting point of 125 °C. A visualisation of the overall intermolecular environment of **12 RN** can be found in Figure 126. Both of the H11 and O1 atoms of the newly introduced aldehyde group are involved in contacts. There is a significant number of p-type interactions present, unsurprising due to the planarity of the molecule within **12 RN**. Finally, the nitro environment in **12 RN** seems crowded, with a large number of contacts.

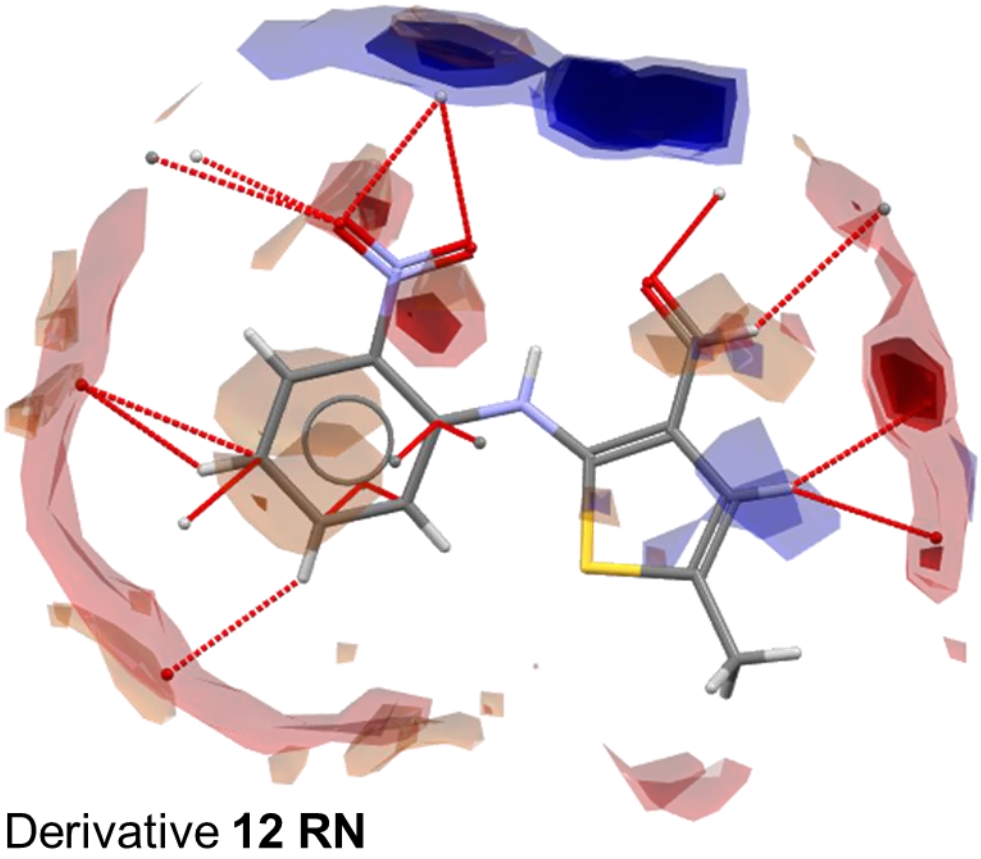


Figure 128; a visualisation of the overall intermolecular environment observed within the crystal lattice of 12 RN.

The contact between the aldehyde H11 proton and the C8 atom at the 4’ position of the nitrobenzene ring is located within the red (carbonyl oxygen) hotspot associated with the aldehyde group, whilst the contact between the aldehyde O1 atom and the H9 proton at the 5’ position of the nitrobenzene ring is located outside of the blue (uncharged NH hydrogen) hotspot associated with the aldehyde group. This suggests a favourable aldehyde environment, though not an ideal one.

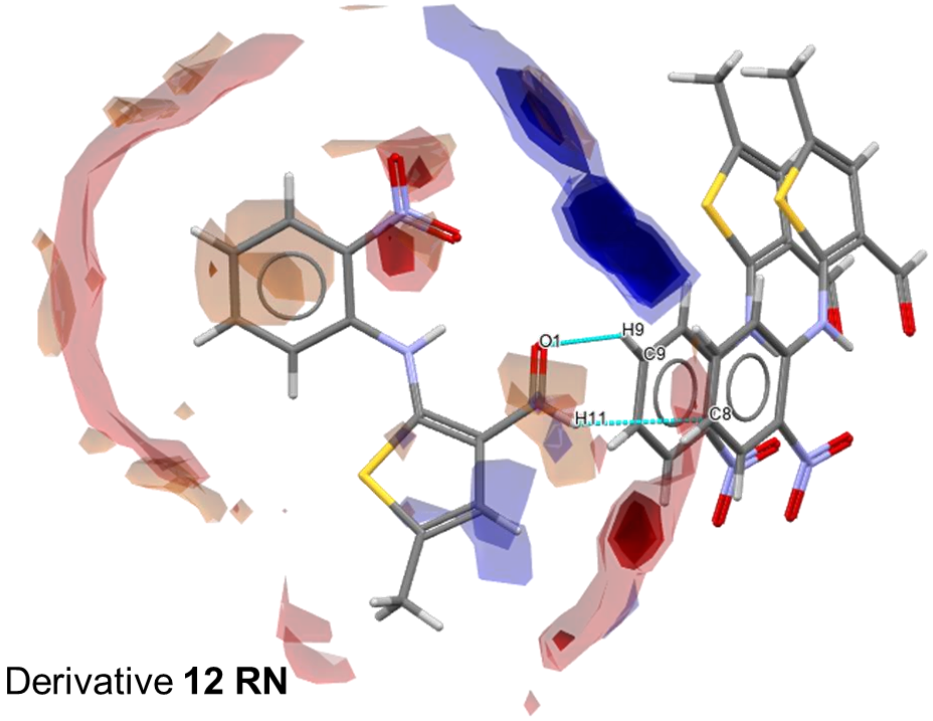


Figure 129; a visualisation of the aldehyde environment within 12 RN

At first glance, the nitro environment in **12 RN** looks crowded and unfavourable, however, expansion of the contacts involving the nitro group reveals that both of the O2 and O3 atoms of the nitro group are involved in contacts with favourable partners. The nitro O2 atom is involved in a contact with the H3 proton at the 4 position of the thiophene ring which is located in the centre of the dark blue (uncharged NH hydrogen) hotspot associated with the nitro group. The nitro O3 atom looks less favourable, displaying two contacts, one with the H8 proton at the 4' position of the nitrobenzene located away from any FIMs hotspots, alongside forming a contact with the H3 proton at the 4 position of the thiophene ring. The O3 has a less favourable environment overall than the O2 atom, however, both nitro oxygen atoms form contacts with favourable partners, and display at least one contact within the dark blue (uncharged NH hydrogen) hotspot, suggesting a favourable nitro environment overall in **12 RN**. A visualisation of the intermolecular environment around the nitro group in **12 RN** can be found in Figure 128.

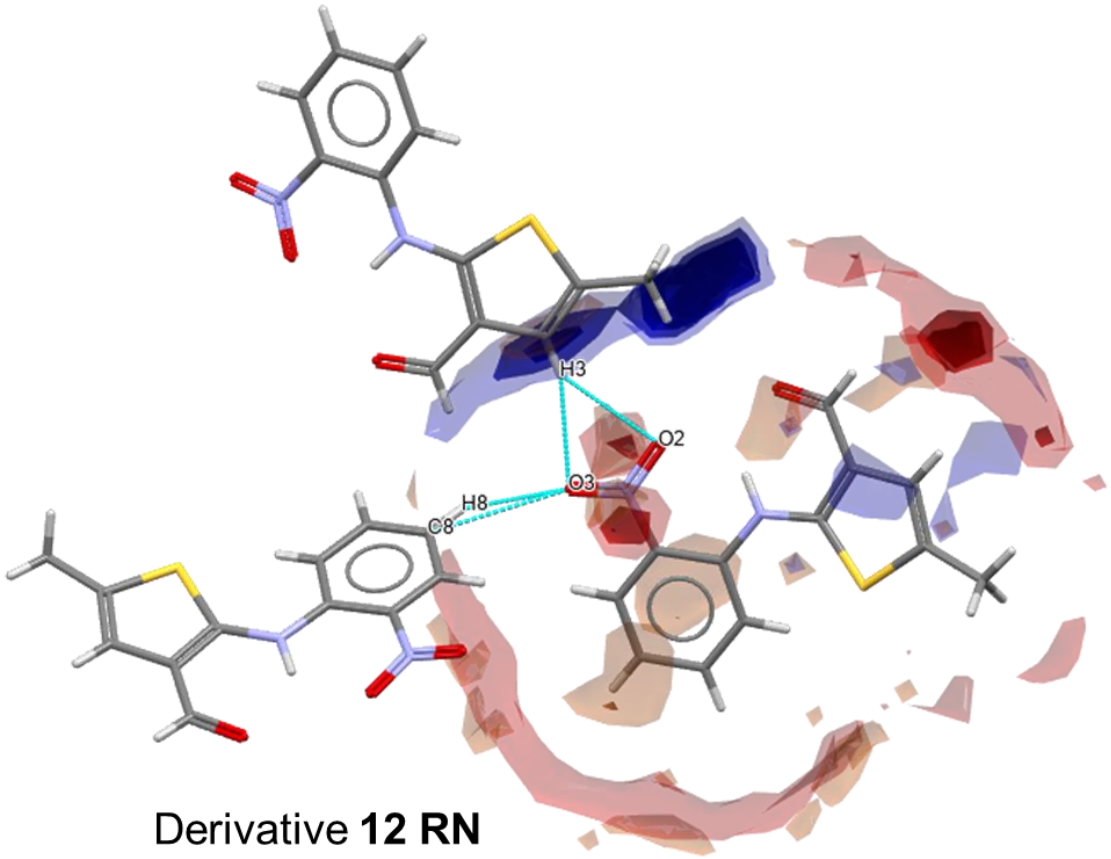


Figure 130; a visualisation of the nitro environment within 12 RN

Overall, the intermolecular environment observed within the crystal lattice of **12 RN** looks favourable, with the major groups involved in favourable contacts, suggesting that the **12 RN** form is a stable form of **12**.

Experiments designed to access crystals of sufficient size and quality for structural determination led to the discovery of an alternate form of **12**, as orange needles (**12 ON**). **12 ON** was crystallised from an excess of chloroform. Unfortunately, no thermal data were acquired for **12 ON** due to time and budget constraint, however, the presence of another structurally characterised form of **12** presents an opportunity to compare the intermolecular environments of **12 RN** and **12 ON** to make an assessment of their favourability. Given the similar intramolecular properties of **12 RN** and **12 ON** (θ_{thio} = 1.33° and 1.67°, respectively), any differences in polymorphic stability can be partially rationalised through analysis of the intermolecular environments observed within **12 RN** and **12 ON**.

The overall intermolecular environment within the crystal lattice of **12 ON** can be found in Figure 129. The overall intermolecular environment for **12 ON** looks surprisingly similar to the environment observed within the crystal lattice of **12 RN**. The main difference at a glance is the lower degree of crowding observed around the nitro group in **12 ON** compared to **12 RN**, along with the absence of

any contacts directly involving the aldehyde H11 proton in **12 ON**. There is little to discern between the two forms from an overview, however, expansion of the contacts around the nitro and aldehyde groups in **12 ON** reveals key differences between the two forms.

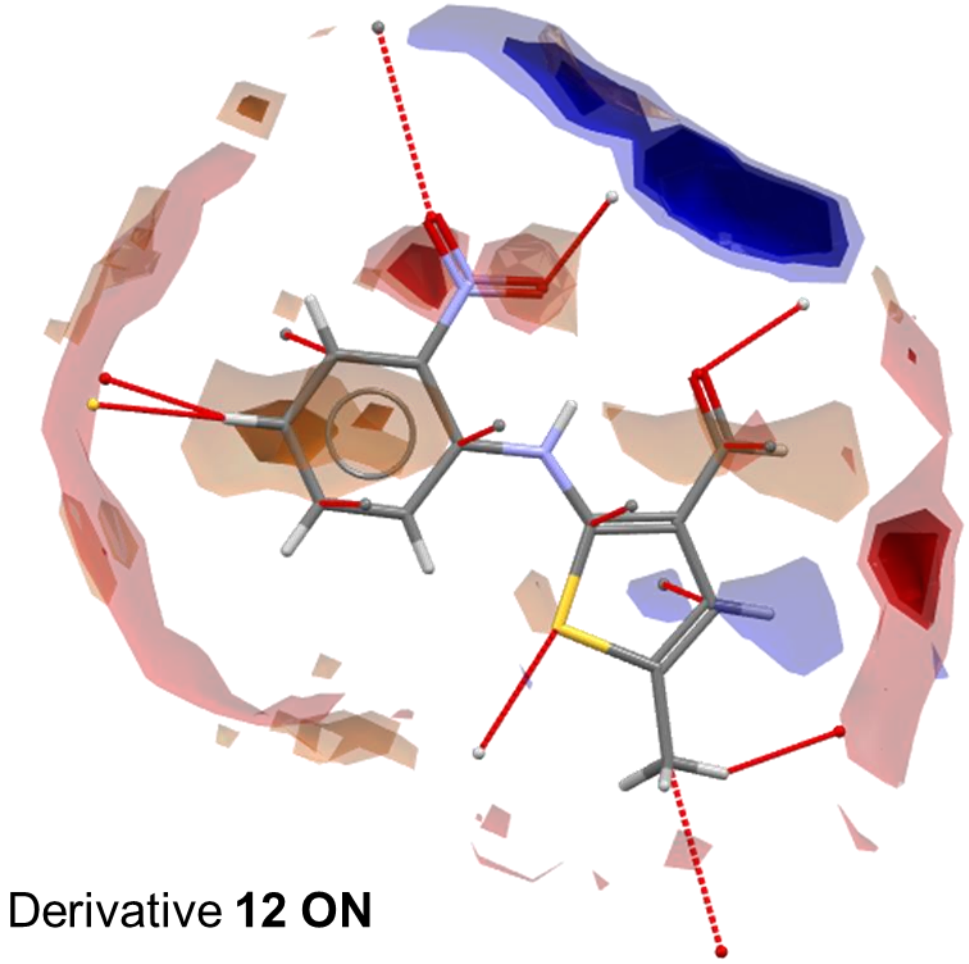


Figure 131; a visualisation of the overall intermolecular environment observed within the crystal lattice of 12 ON.

Figure 130 contains a visualisation of the intermolecular environment of the aldehyde group within the crystal lattice of **12 ON**. Upon closer inspection, the environments of the H11 protons in **12 ON** and **12 RN** are similar, despite the location of the intermolecular contacts. In both cases the interacting partner for the C11/H11 is an aromatic carbon atom, C8 for 12 RN and C9 for **12 ON**. In both of these cases the contacts seem to be a result of a more extended p-system, rather than involving the H11 as a weak H-bond donor. The major difference between **12 ON** and **12 RN** is the environment of the aldehyde O1 atom. In 12 ON the O1 atom is involved in a contact with the thiophene-bound methyl H12B proton, whilst in 12 RN the contact is with the aromatic H9 proton. Since in each form the interaction is located in roughly the same area relative to the blue (uncharged NH hydrogen) hotspot, the favourability of the interaction is largely

down to the compatibility of the partners. Since the H9 proton is more acidic than the H12B proton, it is a more favourable/strong H-bond donor, and hence would form a more favourable interaction with the H0-bond acceptor O1 atom. Therefore, the aldehyde environment in **12 ON** is less favourable than the one observed in **12 RN**.

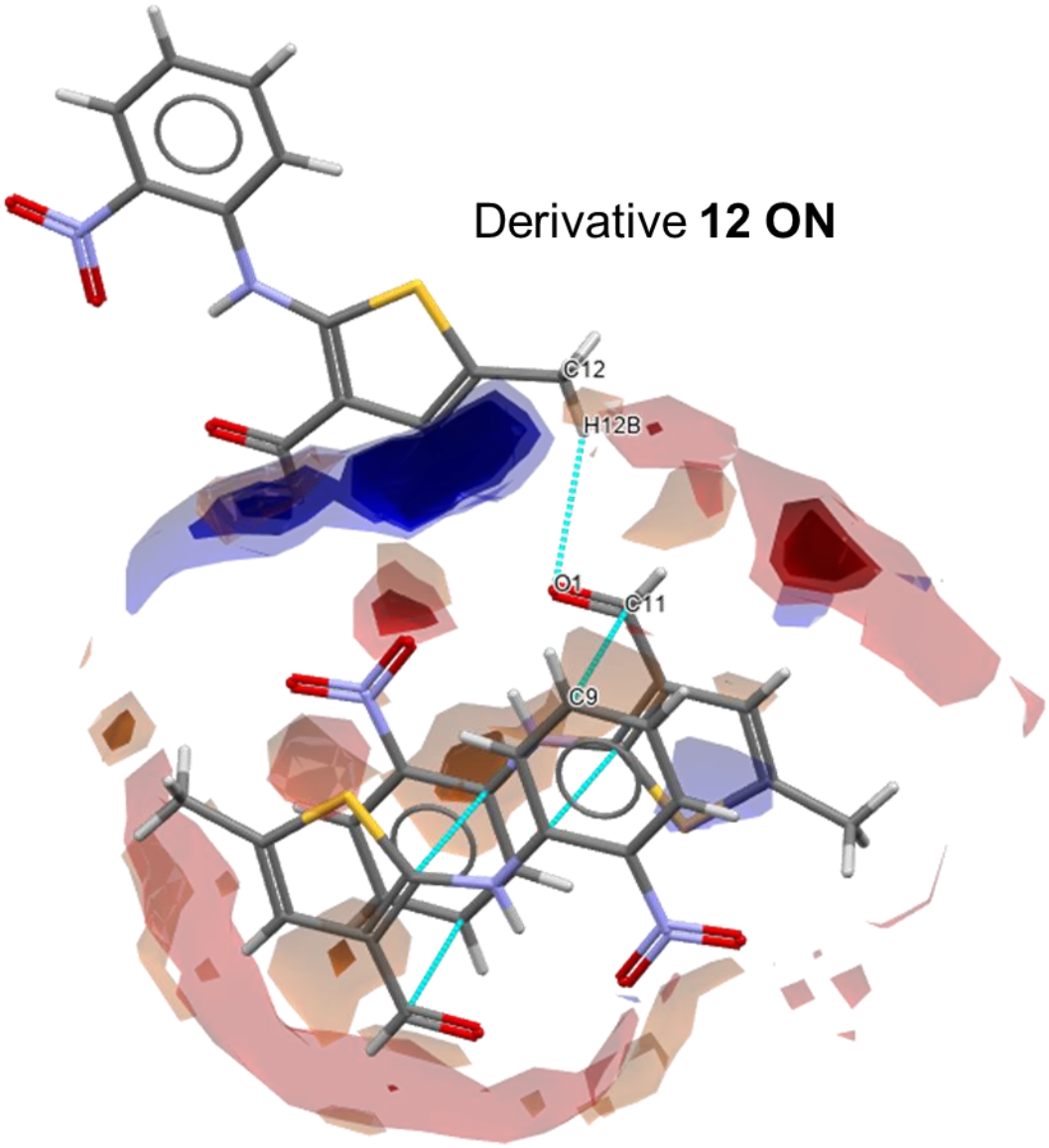


Figure 132; a visualisation of the intermolecular environment of the aldehyde group in 12 ON.

Similarly, the environment of the nitro group in **12 ON** is slightly less favourable than that observed for **12 RN**. The O2 environment is similar, forming a contact ith the aromatic H8 proton within the centre of the dark blue (uncharged NH hydrogen) hotspot associated with the nitro group, suggesting a favourable contact. However, where **12 ON** features a contact with another aromatic proton (see Figure 128) and O3, the O3 atom in **12 ON** forms a contact with the thiophene-bound methyl C12 atom. This results in a less favourable nitro

environment for 12 ON. A visualisation of the nitro environment in **12 ON** can be found in Figure 131.

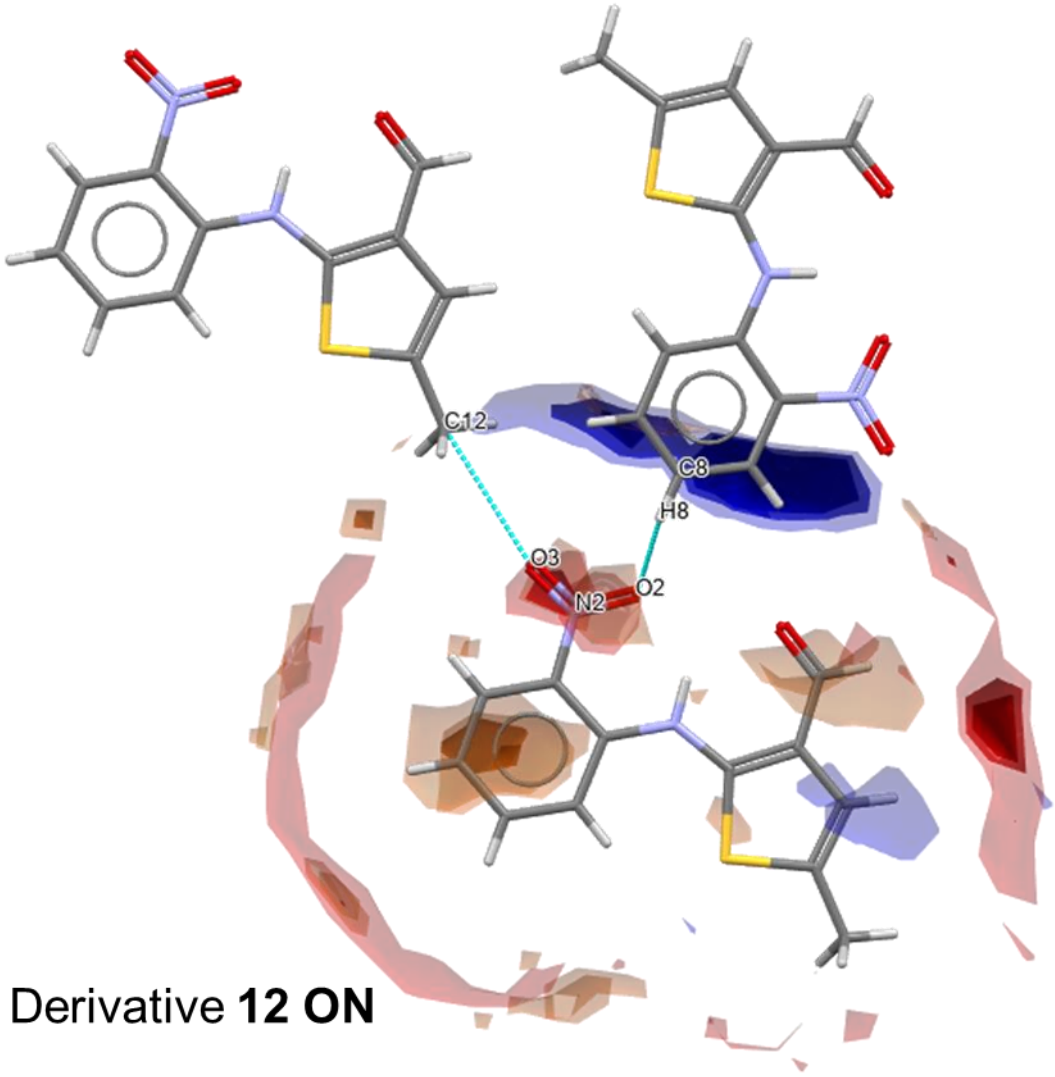


Figure 133; a visualisation of the intermolecular environment of the nitro group in 12 ON.

Analysis and comparison of the intermolecular environments presented within the crystal lattices of **12 ON** and **12 RN** suggests that **12 RN** would be the more stable of the two, however, thermal analysis of **12 ON** would be required to confirm this.

3.6 Supramolecular Environment of ROY Derivatives

In section 2.3 Hydrogen-Bond Propensity (HBP) modelling was used to evaluate the supramolecular environments of the 12 ROY unique polymorphs contained within the Cambridge Structural Database (CSD). Unsurprisingly, due to the presence of an intramolecular H-bond, and the small number of H-bond donors within the ROY structure, this analysis revealed that any H-bond networks formed

within ROY crystals are simple, zero-dimensional (i.e. one donor, one acceptor) networks. Furthermore, most ROY forms (8/12, 75 %) don't display any intermolecular H-bonds. Of the four ROY forms which were found to display intermolecular H-bonds, three are amongst the most stable ROY forms (**Y** (QAXMEH01), **O** (QAXMEH03) and **YT04** (QAXMEH12)), whilst one form, **YN** (QAXMEH04), was found to be metastable. Investigation of the H-bonds observed within these structures revealed that each of the stable forms which were found to display H-bonds contained a CN – HN H-bonds, whilst **YN** (QAXMEH04) contained a NO_2 – HN H-bond, suggesting H-bonds utilising the nitrile H-bond acceptor to be indicative of more stable forms, and those utilising the nitro acceptor group more likely to be metastable.

With the exception of derivative **11**, the modifications made to produce the library of ROY derivatives studied in this work do not alter the number of strong H-bond donors within the scaffold, and hence the observations made for ROY polymorphs in section 2.3 should largely hold true. Using this logic, in this section HBP analysis is performed on the 12 ROY derivatives in order to assess the favourability of the supramolecular environment observed within each. These data are used to assess the stability of the observed form of each ROY derivative, and help to predict whether more stable forms are likely to be uncovered via polymorph screen. Derivative **11** introduces a primary amine group to the ROY scaffold at the expense of the nitro group, significantly altering the donor/acceptor balance relative to the other ROY derivatives, and hence requires additional consideration.

3.6.1 ROY Derivatives: Supramolecular Environment Overview

Derivatives **1** – **10** feature modifications to the ROY scaffold which do not affect the overall balance of H-bond donor/acceptor groups within the ROY scaffold. This means that the potential H-bonding networks which can be formed within crystal structures of derivatives **1** – **10** are formed with the same donor/acceptor partners as the ROY polymorphs discussed in 2.3. The only differences in the propensity and coordination scores for **1** – **10** are a result of the influence of the group (methyl or chloro) introduced to the ROY scaffold, and how this group would affect the steric or electronic environment of the donor/acceptor groups.

Derivative **12 RN** and **12 ON** are slightly different, in that the nitrile group present in the ROY scaffold is replaced with an aldehyde group. As the aldehyde group is also a strong H-bond acceptor, the balance of H-bond donor/acceptor groups in forms of 12 is the same as with the ROY scaffold. However, the aldehyde group does influence the potential H-bonding networks via the introduction of a second intramolecular H-bond involving the bridging amine proton. This intramolecular

H-bond further reduces the likelihood of the formation of intermolecular H-bonds relative to the ROY forms discussed in 2.3. An image describing the predicted and observed H-bonding networks for derivatives **1** – **10**, **12 RN** and **12 ON** can be found in Figure 132.

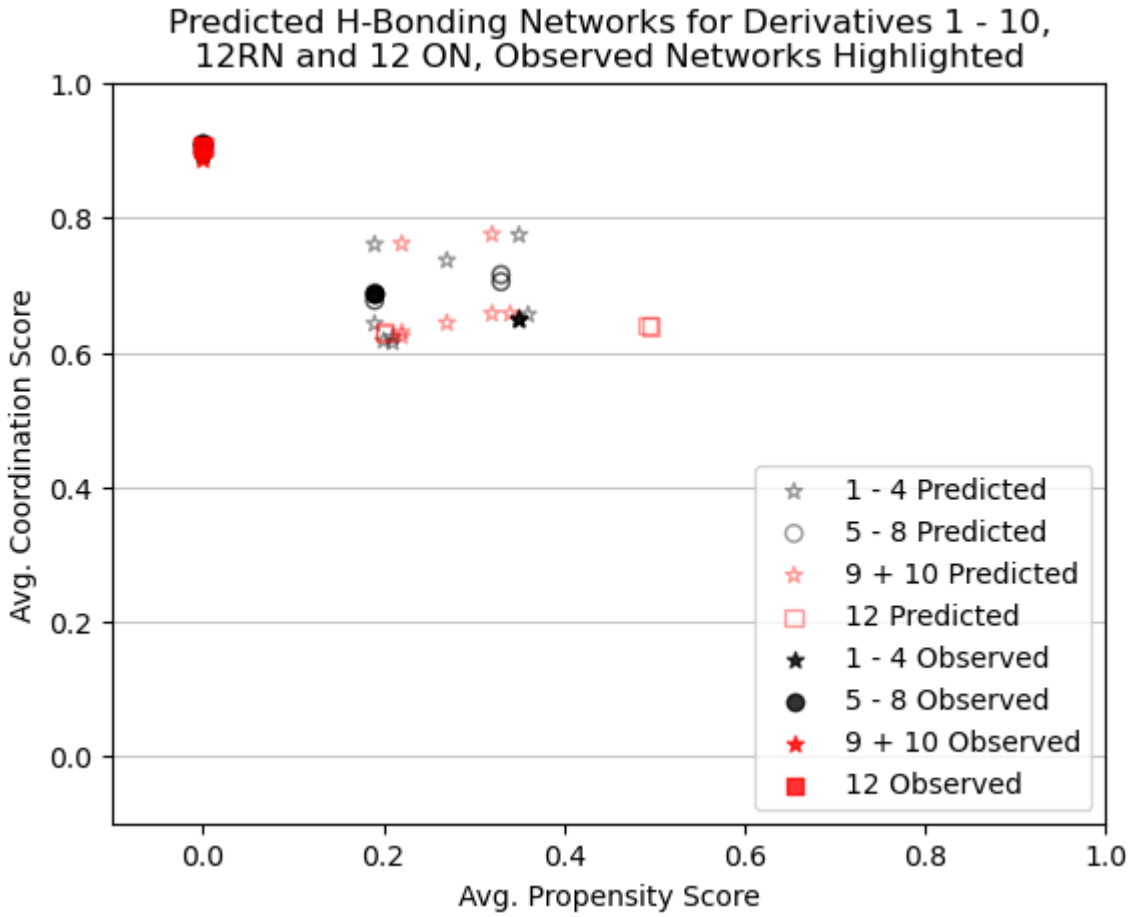


Figure 134; scatterplot of average coordination score against average propensity score of potential H-bonding networks contained within derivatives 1 – 10, 12 RN and 12 ON. ROY Derivatives for which modifications were made to the nitrobenzene ring are highlighted in black, aminothiophene-based modifications are highlighted in red. Methyl-modifications are denoted by stars, chloro-modifications by circles and aldehydes by squares. Predicted H-bonding networks are denoted by unfilled ticks, observed networks are denoted by filled ticks.

The similarities between the distribution of predicted and observed H-bonding networks for the ROY polymorphs discussed in 2.3 and that of derivatives **1** – **10**, **12 RN** and **12 ON** is immediately apparent. Amongst the derivatives discussed above, only three were found to display intermolecular H-bonds: **3 RN** (MP = 145 °C), **5 ON** (MP = 135 °C) and 7 YP (MP = 155 °C). Without thermal data for other polymorphic forms of 3, 5 and 7, it is impossible to rationalise the effect of intermolecular H-bonds on polymorphic stability. However, investigating which H-bonds form could provide insight into the influence of the modification on favoured H-bonding motifs.

3 RN is the only form found to display an intermolecular H-bond between the nitrile nitrogen and the bridging amine proton, which was the most prominent H-bond observed amongst the ROY polymorphs discussed in 2.3. 3 RP has the highest melting point amongst derivatives **1** – **4**, which could suggest a link between the presence of H-bonds and polymorphic/crystal stability, however, additional polymorphic data for **1** – **4** would be necessary to confirm this. An image of the CN – HN H-bond within the crystal lattice of 3 RN can be found in Figure 133.

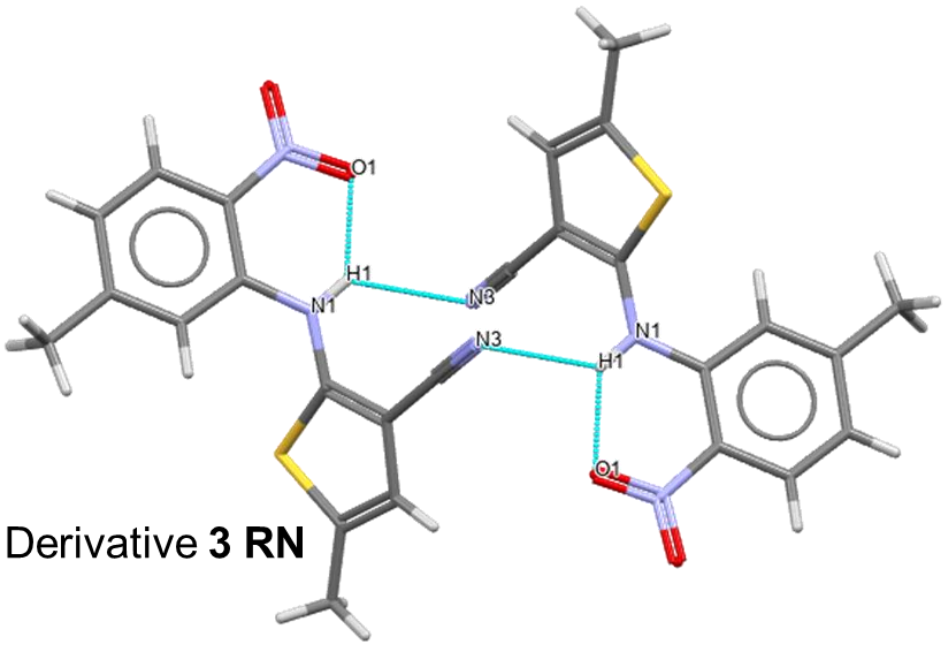


Figure 135; visualisation of the CN – NH intermolecular H-bond observed within the crystal lattice of 3 RN.

Both derivative **5 ON** and **7 YP** are found to display an intermolecular H-bond between the nitro and bridging amine groups. In **5 ON** it is the nitro O2 atom, which is not involved in an intramolecular H-bond, which forms the contact with the bridging amine H1 proton, whereas in **7 YP**, it is the nitro O1 atom which forms the intermolecular H-bond. This results in H-bonding motifs which look surprisingly different, as highlighted in Figure 134. As mentioned previously, the lack of thermal/polymorph data for derivatives **5** – **8** makes evaluation of the impact of NO_2 – NH H-bonds on polymorphic stability moot, however, the presence of two NO_2 – NH H-bonds amongst **5** – **8** could suggest that the introduction of a chlorine atom around the nitrobenzene ring in ROY could favour the formation of NO_2- NH intermolecular H-bonds.

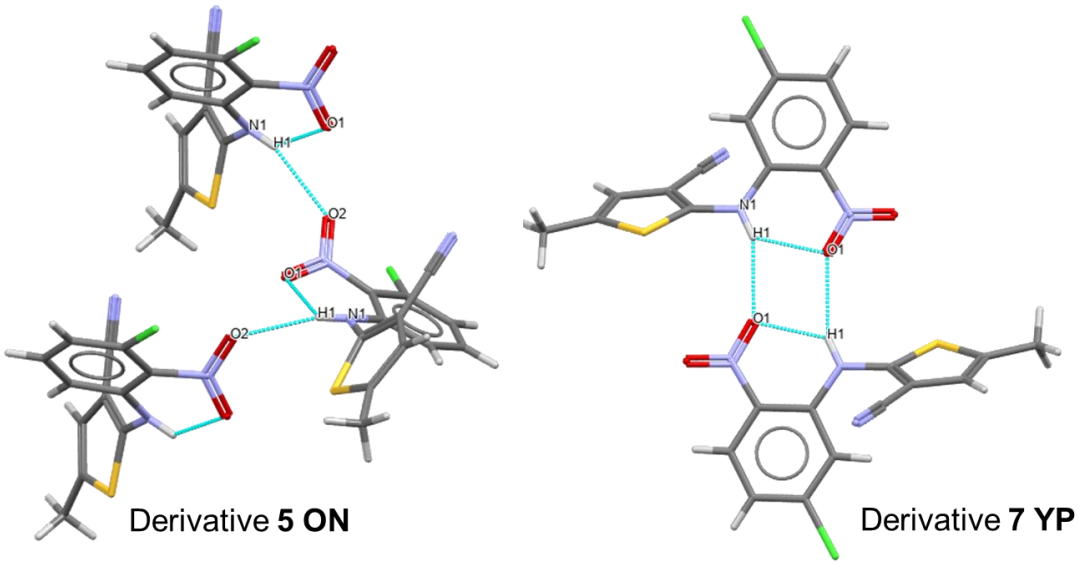


Figure 136; visualisation of the NO_2 (O2) – NH intermolecular H-bond present within the crystal lattice of 5 ON (left), and the NO_2 (O1) – NH intermolecular H-bond present within the crystal lattice of 7 YP (right).

3.6.2 Derivative 11 CB (XABXEF) Supramolecular Environment

In converting the ROY scaffold to derivative **11** two H-bond acceptor atoms (nitro O1 and O2 atoms) are replaced by two H-bond donors (amine H1 and H2 protons) and one H-bond acceptor atom (primary amine N2 atom). This alteration significantly shifts the balance between donor and acceptor atoms within the scaffold of **11** relative to ROY, and would likely encourage the formation of intermolecular H-bonds (due, in part, to the removal of the intramolecular H-bonds between the bridging amine and nitro groups). This results in a significantly more complex HBP model for **11**, in which networks consisting of one, two and three discrete H-bonds are possible. Figure 135 contains a scatterplot describing the average propensity score and coordination scores calculated for the H-bonding pairs within the scaffold of **11**, sorted by their number of constituent H-bond pairs. The H-bond network observed within the crystal lattice of **11 CB** is highlighted.

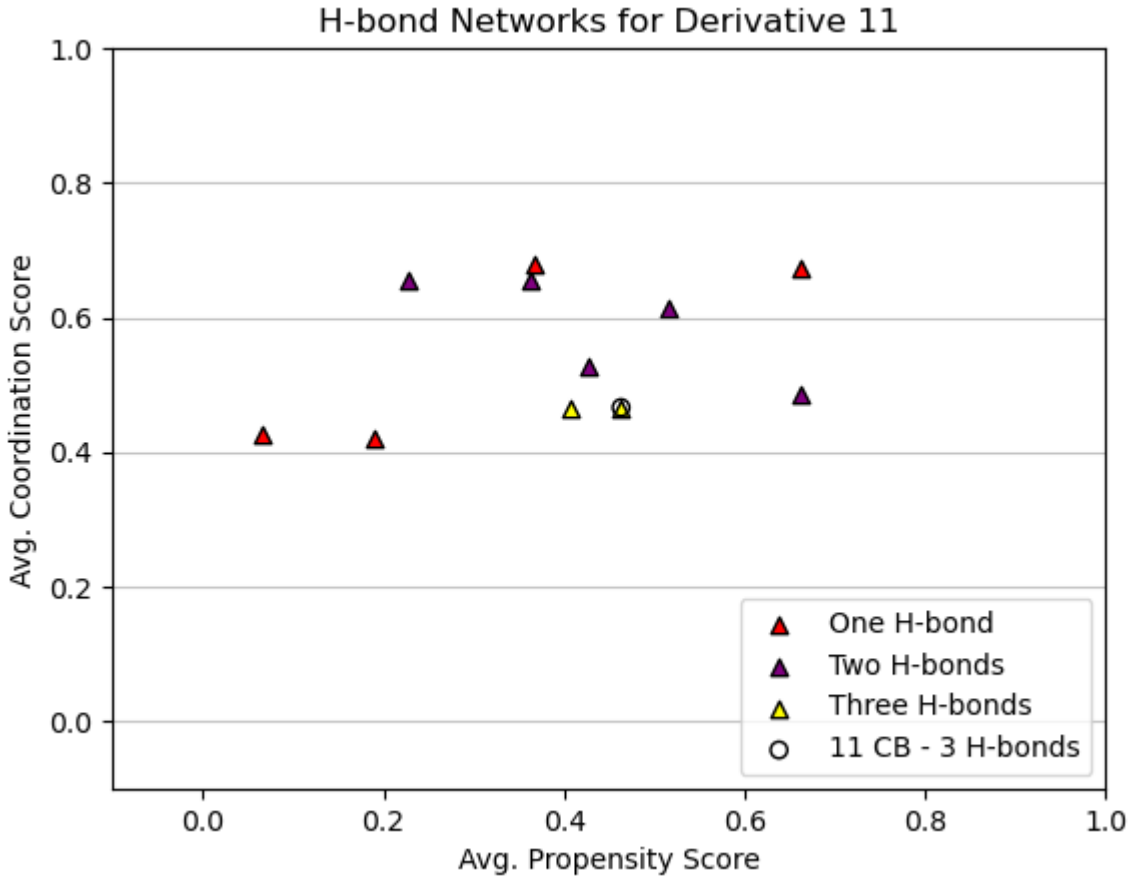
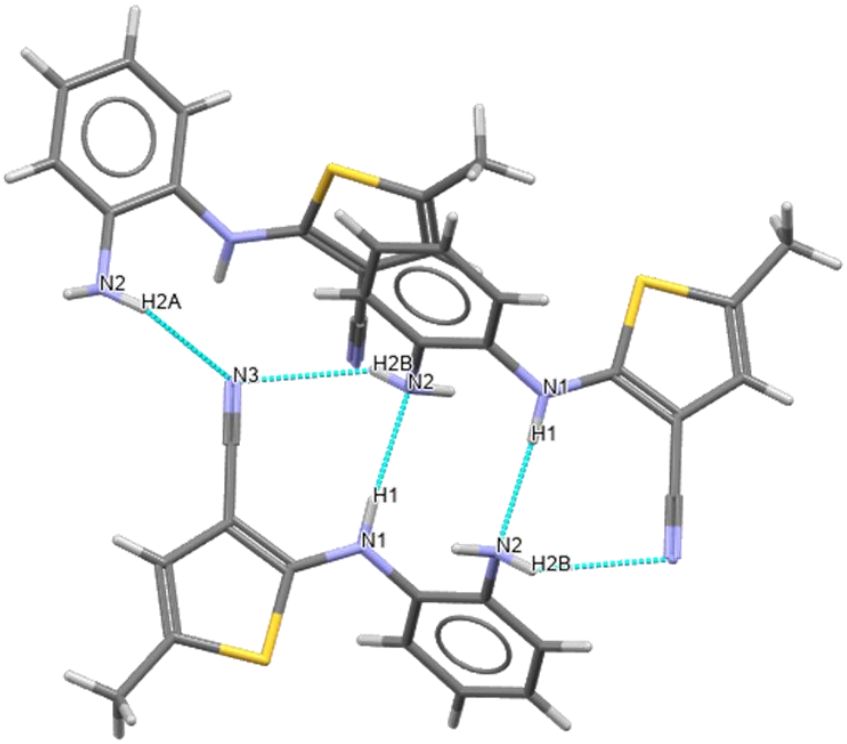


Figure 137; scatterplot of average coordination score against average propensity score of potential H-bonding networks contained within derivative 11 CB/XABXEF. H-bond networks consisting on one, two and three intermolecular H-bonds are denoted by red, purple and yellow triangles, respectively. The H-bonding network observed within the crystal lattice of 11 CB/XABXEF is denoted by a white circle.

As highlighted in 3.3.4, within the crystal lattice of **11 CB** there are three discrete H-bonds: two between the primary amine protons and the nitrile N3 atom, and another between the bridging amine proton and the nitrogen atom of the primary amine. The average propensity and coordination scores predicted by the HBP model for derivative **11** place this network towards the bottom of the scatter graph, suggesting a less favourable H-bonding network relative to other predicted networks. This is likely due to the presence of the two H-bonds centred on the nitrile N3 atom, resulting in a bifurcated acceptor atom, a term which the HBP model is biased against. This conflicts with the observations regarding polymorphic stability made in section 3.3.4 and with the fact no other forms of **11 CB**/XABXEF have been discovered since its first characterisation by Leyva-Perez et al. A visualisation of the H-bonds observed within the crystal lattice of **11 CB**/XABXEF can be found in Figure 136.



Derivative **11 CB**

Figure 138; a visualisation of the NH2 (H2A) – CN, NH2 (H2B) – CN and NH – NH2 intermolecular H-bonds observed within the crystal lattice of 11 CB/XABXEF.

One potential explanation for the discrepancy in the predicted favourability of the H-bond network observed within the lattice of **11 CB** is the aforementioned bias against bifurcation in the HBP model. This was also observed for ROY derivatives discussed in 2.3, which saw CN – NH H-bonds as having low propensity and coordination scores, despite thermal/stability data suggesting the presence of CN – NH H-bonds to be indicative of more stable ROY forms. This could suggest that the models used to calculate H-bond coordination and H-bond propensity are too strongly biased against the presence bifurcated atoms. This may be amenable with the model used to calculated H-bond coordination, which is a static model. However the propensity model is constructed from data across the CSD, and any bias against bifurcated atoms is resultant from the trends across the CSD.

Another explanation for the discrepancy could be the result of the chain-like motif formed by H-bonds in **11 CB**. This could theoretically add further long range stabilising effects within the **11 CB** crystal lattice, effects which may not be captured by the propensity or coordination models at this stage. Derivative **11 CB** serves as a very interesting case study into the predictive power of HBP and SFI as a whole, and further experiments, potentially involving the discovery of polymorphs of **11**, would be very useful.

3.7 ROY Analogues Polymorph Screen

One of the main goals of this work was to explore the position of the ROY scaffold as a potential polymorphophore, one of the key features of which is the ability to produce structural analogues which display polymorphism. In order to explore this, a simple, low-/medium-throughput polymorph screen was designed, implemented and applied to each of the 12 ROY analogues synthesised for this work.

Typically, polymorph screens are performed industrially with a view to identifying the ideal candidate form to be taken further into development. This typically entails utilising a range of crystallisation techniques designed to thoroughly explore the polymorphic landscape of the candidate compound, with the goal of identifying a crystalline form with properties most suited for the end task of the product. These techniques are used in conjunction with characterisation techniques which can identify different crystalline forms such as PXRD, thermal analysis or Raman spectroscopy. A summary of some techniques utilised in industrial polymorph screens can be found in Chapter 1.

As the goals of this work are to confirm the presence of polymorphism in the 12 ROY analogue compounds, the polymorph screen utilised in this work is significantly simplified when compared to those carried out industrially. Slurrying experiments, for instance, can take time on the order of weeks to months, and are typically utilised to identify thermodynamically stable forms. However, the timescale of this project, along with the indifference towards the specific properties of the observed polymorphic forms, meant that longer timescale experiments were not necessary/considered for this work.

This section will discuss the logic behind the design of the polymorph screen implemented in this work, followed by a discussion of the findings of the polymorph screen itself. Finally, the implications of these data will be discussed in regards to the position of the ROY scaffold as a potential polymorphophore.

3.7.1 Design of the low-medium-throughput Polymorph Screen

Before the screen could take place there were two key limiting factors which would come to define the number and nature of any possible experiments: time and availability of material. The limited time available, along with the number of compounds to screen, meant that any screening experiments performed must be achievable on a short timescale (on the order of hours to days). This inherently ruled out longer scale experiments such as slurry or seeded experiments, along with any techniques utilising specialist equipment which would require training/hiring of equipment, such as grinding/liquid assisted grinding

experiments. With these limitations in mind, the experiments chosen to make up the screen were evaporation crystallisation and antisolvent crystallisation.

Utilising a range of solvents in evaporation experiments allowed for variation in the relative supersaturation across said experiments due to differences in solubility in each solvent, whilst modifications to the evaporation rate of these experiments would lead to further differences in supersaturation, and hence the possibility of accessing different polymorphic forms. The addition of antisolvent to the different evaporation experiments would also have a similar effect.

The amount of material available was the other limiting parameter when designing the polymorph screen. The amount of material available for each compound varied between ~ 750 mg – 2000 mg, thus a maximum of 500 mg of material was allocated for screening purposes, in order to avoid depletion of stocks. The goal when designing the screen was to maximise the number of experiments performed, and hence chances at observing polymorphism, whilst minimising the amount of material used.

The screen was to be performed using a 4x8 array of experiments per compound, in which 2 mL LC-MS vials were used as crystallisation vessels.

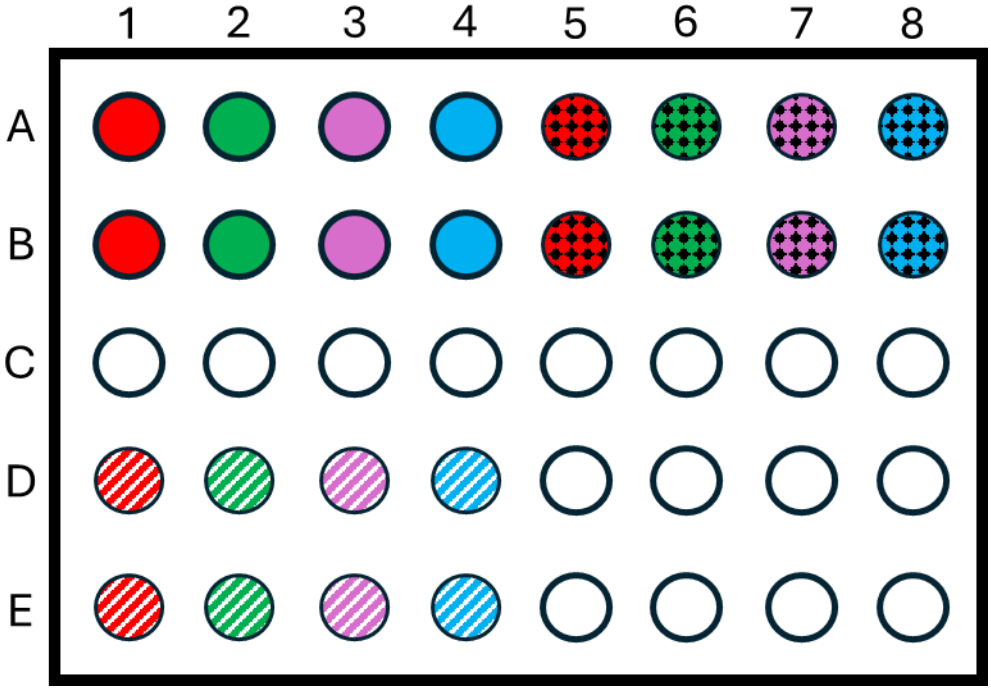


Figure 139; each well/cell represents a 2 mL vial in which a ROY analogue is subject to conditions designed to access alternate polymorphic forms. All experiments are performed at room temperature on 10 mg of ROY analogue, using a total volume of 2 mL of the specified solvent and subject to the following conditions: (Row A) column 1 = acetone, column 2 = acetonitrile, column 3 = chloroform, column 4 = ethyl acetate, column 5 = A1 covered with aluminium foil, column 6 = A2 covered with aluminium foil , column 7 = A3 covered with aluminium foil, column 8 = A4 covered with aluminium foil. (Row B) Repeat of Row A. (Row C) Blank. (Row D) Column 1 = 10:1 acetone/water, Column 2 = 10:1 acetonitrile/water, Column 3 = 10:1 chloroform/ethanol, Column 4 = 10:1 ethyl acetate/hexane. (Row E) Repeat of Row D.

A key factor to consider when designing a polymorph screen is how to detect instances of polymorphism. Different crystal forms of the same compound display different physicochemical properties, thus detection of different polymorphic forms can be achieved through examination of these properties relative to those of the form characterised during Chapter 3.

The ROY scaffold is unique in that polymorphs of ROY are unusually susceptible to colour polymorphism, in which different ROY polymorphs can display distinctly different colours, alongside the expected differences in physicochemical properties. This effect makes the identification of different ROY polymorphs much easier than for other scaffolds.

During the synthesis and characterisation of ROY analogues in Chapter 3, it was noted that these derivatives crystallise in brightly coloured forms similar to those observed for the ROY polymorphs. Moreover, it was found that some of these

analogues were susceptible to undergo colour polymorphism, as highlighted in Figure 140.

Figure 140; three polymorphic forms of compound 4 identified during the polymorph screen. The red form on the left was identified as 4 R, whilst the middle and rightmost forms were identified as new polymorphic forms of 4.

It was decided that the colour polymorphism of ROY analogues could be utilised as a rapid and simple detection method for the presence of new polymorphic forms. Figure 141 contains PXRD patterns of compound 4 R and an orange needle form of 4 identified during the polymorph screen, along with their corresponding crystal.

Powder X-Ray Diffraction (PXRD) is a rapid and simple structural characterisation technique which allows for the identification of prominent faces within a given crystal lattice through sweeping an X-Ray source through a powdered sample of crystalline material across a range of angles. This produces a characteristic pattern of peaks in the X-Ray absorption spectrum which can be used to identify specific crystals of a given material. Through comparison of PXRD patterns generated for potential polymorphs discovered via the polymorph screen with predicted PXRD patterns taken from the CIF of forms characterised in Chapter 3, coupled with the observation of crystal forms of different colour and/or habit from those forms characterised in Chapter 3, new polymorphic forms of the ROY analogues can be identified.

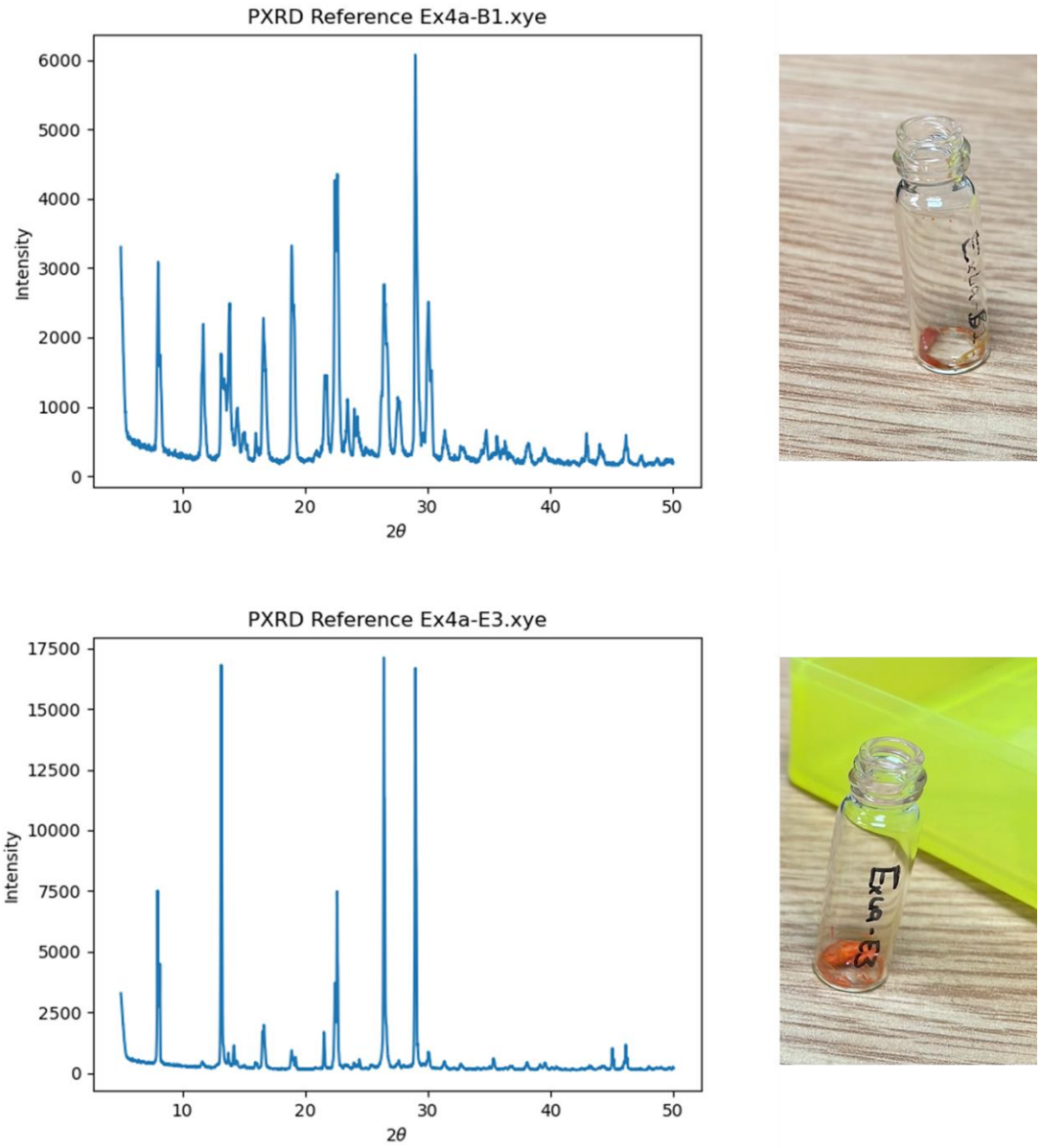


Figure 141; (Top) Red prisms of compound 4 identified as 4 R, next to the corresponding experimental PXRD pattern. (Bottom) Orange needles of compound 4 next to the corresponding experimental PXRD pattern. Comparison of this PXRD pattern to the simulated pattern of 4 R indicates that this is an undiscovered polymorph of 4.

3.7.2 Polymorph Screen Results

Throughout the synthesis and characterisation of the 12 ROY derivatives it became apparent that derivatives 1 – 8 crystallised in distinctive, brightly coloured crystals similar to those observed for the different ROY polymorphs, whereas derivatives 9 – 12 formed less brightly coloured forms. This observation also held for the results of the polymorph screen, with 1 – 8 readily forming new brightly coloured forms, making the identification of potential new forms quick and simple.

A table summarising the results of the low-/medium-throughput polymorph screen for compounds **1** – **8** can be found in Table 13

Despite its relative simplicity, the polymorph screen was successful in identifying alternate polymorphic forms for all eight of the ROY analogues featuring additions to the nitrobenzene ring. The potential new forms were easily identified due to their change in colour/form relative to the seed/input form used at the beginning of the screen. In order to confirm the presence of a new crystal form, PXRD patterns were taken of these distinctive crystals, which was then compared using a python script provided by the CCDC to the predicted pattern of the input/seed form to ensure a new crystal form was present.

Table 13; summarising the results of the polymorph screening experiments performed on compounds 1 – 8. Potential and confirmed polymorphic forms are highlighted along with their crystal habit, colour, PXRD pattern match with seed form and the experimental conditions used to access each polymorph.

Reference Form	Conditions	Colour	Habit	PXRD (%)	Polymorph Evaluation
1 ON	Acetone	Yellow	Prism	87.57	Undiscovered polymorph
1 ON	10:1 Acetone/Water	Yellow	n/a	90.57	Inconclusive
2 R	Acetone	Orange	Needle	79.95	Undiscovered polymorph
2 R	Ethyl Acetate (Covered)	Orange	Needle	87.05	Undiscovered polymorph
3 RN	Acetone	Orange	Needle	87.57	Undiscovered polymorph
3 RN	10:1 acetone/Water	Orange	Needle	88.52	Undiscovered polymorph
3 RN	Acetonitrile (Covered)	Orange	Needle	88.76	Undiscovered polymorph
4 R	Acetone	Red	Prism	95.38	**4 R**
4 R	10:1 Acetone/Water	Orange	Needle	93.08	Undiscovered polymorph

5 ON	Acetone (Covered)	Red	Prism	81.75	Undiscovered polymorph
5 ON	Chloroform (Covered)	Red	Prism	82.89	Undiscovered polymorph
6 RN	Acetone	Yellow	Prism	91.82	Undiscovered polymorph
6 RN	Acetonitrile (Covered)	Yellow	Prism	87.76	Undiscovered polymorph
7 Y	Acetone	Red	Prism	88.22	Undiscovered polymorph
8 RN	10:1 Ethyl Acetate/Hexane	Orange	Needle	91.53	Undiscovered polymorph

In contrast to **1** – **8**, crystals of compounds **9, 10** and **12** were found to form dark red crystals, significantly darker than any of the ROY polymorphs discovered to date, whilst crystals of compound **11** were found to be colourless. The polymorph screen of compounds **9** – **12** revealed no significant observable change in colour, making the visual detection of potential new polymorphs much more difficult. For instance, during characterisation of the ROY analogues, two polymorphs of compound **12** were identified, one a bright red needle (**12 RN**) and one more pale-red/orange needle (**12 ON**), owing to the similar θ_{thio} angle observed for the two forms. Both compound 9 and compound 10 were found to display similar θ_{thio} angles to **12 RN** and **12 ON**, adopting a nearly planar. As a result of this, identification of potential polymorphs of **9** – **12** was significantly more difficult, with judgement of the presence of new polymorphs done largely via visual comparison of crystal habit, rather than the combination of differences in colour and/or habit.

Potential new polymorphic forms of **9** – **12** were identified, with PXRD pattern comparison/analysis suggesting new crystalline forms of **9** – **12** have been identified. However, the results are inconclusive, and further data are required to confirm the nature of the potential new polymorphs of **9** – **12**. A table summarising the results of the low-/medium-throughput polymorph screen for compounds **9** – **12** can be found in Table 14.

Table 14; summarising the results of the polymorph screening experiments performed on compounds 9 – 12. Potential and confirmed polymorphic forms are highlighted along with their crystal habit, colour, PXRD pattern match with seed form and the experimental conditions used to access each polymorph.

Reference Form	Conditions	Colour	Habit	PXRD Match (%)	Polymorph Evaluation
9 DR	10:1 Acetone/Water	Red	Prism	91.07	**9 DR**
9 DR	10:1 Ethyl Acetate/Hexane	Red	Prism	82.14	Potential Polymorph
10 R	Acetone	Red	Prism	91.19	**10 R**
10 R	Chloroform	Red	N/a	81.06	Potential Polymorph
11 CB	Chloroform (Covered)	Colourless	N/a	88.45	Inconclusive (Low crystallinity)
11 CB	10:1 Acetone/Water	Colourless	N/a	86.14	Inconclusive (Low crystallinity)
12 RN	Chloroform	Red/Orange	Needle	80.46	Potential Polymorph
12 RN	10:1 Acetone/Water	Red/Orange	Needle	82.51	Potential Polymorph
12 ON	Chloroform	Red/Orange	Needle	88.69	**12 ON**
12 ON	10:1 Acetone/Water	Red/Orange	Needle	88.93	Potential Polymorph

Despite the inconclusive nature of the data, comparison of PXRD patterns of characterised forms of **9** – 1**2** with selected polymorph screen experiments revealed that both **9** and **10** could potentially be polymorphic, whilst a third polymorph of **12** may also be present. In total, this means that the low-/medium-throughput polymorph screen found 8/12 of the ROY derivatives synthesised for this work to display polymorphism, with a further two derivatives (**9** and **10**) potentially displaying polymorphism. With a very simple polymorph screen revealing between 75% – 83% of the ROY derivatives synthesised for this study

to display polymorphism, the data suggest that the ROY scaffold is a polymorphophore.

3.7.3 Polymorph Screen Discussion

In this section, the 12 ROY analogues synthesised and characterised in Chapter 3 were subject to a low-/medium-throughput polymorph screen. Each compound was subject to 12 unique crystallisation conditions and their repeats, for a total of 24 experiments per compound, and 288 experiments overall. New polymorphic forms of compounds **1** – **8** and **12** were observed, whilst potential new polymorphic forms of **9** and **10** were observed. These data confirm that 9 of the 12 ROY analogues (83%) display polymorphism, with the possibility of a further two. Using the definition of polymorphophore defined by Matzger in 2008, these data would suggest that the ROY scaffold is an example of a polymorphophore.

However, despite the overall success of the polymorph screen, a number of issues with the experimental setup were identified which may be improved upon in any subsequent screens. Firstly, whilst observation of a change in colour and/or habit of ROY derivatives was a quick and easy method of identifying new polymorphs, colour polymorphism was only reliably observed for compounds **1** – **8**, with more subtle changes observed for **12**. Using the human eye to judge subtle changes in colour and/or morphology is an unreliable method of identifying new polymorphs. A more objective method of detection such as Raman spectroscopy or thermal analysis/MP would be a more accurate and reliable method of detecting new polymorphic forms.

There were also a number of issues related to the detection of polymorphic forms via PXRD. Typically, when comparing PXRD patterns of structures for a match in polymorphic form, a threshold of > 95% pattern match is used to determine whether two forms are identical. However, in this work a 95% structural match was only observed once, with typical values of up to 93% being observed for identical forms. A key contributing factor to these low scores may be the script used for PXRD comparison, provided by the CCDC, utilises the predicted powder pattern from the CIF of the reference structure to compare with the experimental powder pattern observed from the screening experiments. A direct comparison between experimental patterns would provide a more reliable comparison of structures, and hence identification of polymorphic forms.

Another contributing factor to the low PXRD pattern matches could also be due to the amount of material used in each screening experiment. 10 mg of each compound was used per experiment, however analysis of the PXRD patterns observed from the screening experiments revealed some patterns to be poorly defined. This may be a result of low crystallinity or preferred orientation effects.

Using a larger amount of material per experiment should lead to more well defined powder patterns, and comparing these against an experimental powder pattern for the reference form should provide more reliable comparisons between the reference form and any forms obtained via the polymorph screen, and hence more robust polymorph detection.

Chapter 4 Conclusions and Future Work

The focus of this chapter is to summarise the findings of chapters 2 and 3, and to provide suggestions for further investigations which could expand on the findings of this work. The first section in this chapter is focused on evaluating the efficacy of the SFI workflow in evaluating the polymorphic landscape of solid forms, the ability of SFI to de-risk forms during a screening exercise, and what improvements could be made to the workflow in the future. The second section in this chapter is focused on the structure-polymorphism relationship (SPR) of ROY, and evaluating the status of ROY as a polymorphophore. The results of the SFI analysis performed on ROY derivatives in Chapter 3 is discussed, and what these data suggest about the SPR of ROY. The results of the polymorph screen are discussed, including the efficacy of the screen and the implications the polymorph screen data have on the classification of ROY as a polymorphophore.

4.1 Solid-Form Informatics

The other main objective of this work is to evaluate the efficacy of the SFI workflow laid out first by Galek et al.[38], utilising the most up to date implementation of the informatics tools available for use with the CSD. The SFI workflow uses the informatics tools available within Mercury to explore the intramolecular, intermolecular and supramolecular environments presented by an input structure, in order to explore the polymorphic landscape of the input structure.

In this work, the intra-, inter-, and supramolecular environments of a set of 12 ROY polymorphs and 12 ROY derivatives were explored using Mogul Geometry Check (MGC), Full Interaction Maps (FIMs) and Hydrogen-Bond Propensity (HBP) modelling, respectively, with a view to exploring the structure-polymorphism relationship (SPR) for thee ROY polymorphophore. This section is focused on evaluating the efficacy of the tools contained within the SFI workflow at exploring the polymorphic landscape of the ROY polymorphophore, and to provide suggestions for how the workflow could potentially be improved.

4.1.1 Mogul Geometry Check

MGC is an informatics tool within Mercury which is used to compare the values of certain intramolecular features (bond lengths, bond angles, intramolecular torsions and ring geometries) to examples of similar molecular functionality across the CSD. MGC provides rapid statistical analysis of these intramolecular features, allowing features which do not follow the trends of the library of

structures containing similar functionality constructed by MGC to be highlighted as *unusual*.

This feature is highly useful to the SFI workflow, as unusual geometries may be indicative of unfavourable/metastable polymorphic forms, however, there are a number of key issues which restrict the efficacy of MGC when looking to explore the polymorphic landscape of a given structure

The first issue relates to the ability of MGC to handle degenerate structural representations of input structures. When performing MGC analysis of ROY polymorphs in section 2.1, MGC search of bond lengths for the **ON** (QAXMEH) form resulted in 7 hits for the C-N-C-S (θ_{thio}) torsion, whilst the same search for the **Y04** (QAXMEH53) form yielded 114 hits. The discrepancy in search results is due to the different representations of the thiophene ring in ON and Y04, as ON features aromatised double bonds, whilst **Y04** displays explicit (kekulised) double bonds. This results in two incomplete libraries of structural related structures to the ROY polymorphs, which must be combined manually to produce a more complete library of structures containing structurally similar features to ROY across the CSD. This issue could be circumvented within MGC via standardisation of input structures before performing a structure search of the CSD. Or alternatively, multiple searches for degenerate structural representations could be performed at the start when using MGC as a part of the SFI workflow.

The second issue relates to the issue of duplicate data within the libraries constructed via MGC. MGC has no means to determine whether to structures within its dataset are identical (resulting from structural redeterminations etc.), which can result in the presence of duplicate structures within comparison libraries and hence statistical bias towards those duplicate features. This is an issue which has seemingly been addressed already by Cruz-Cabeza et al. in their 2015 paper “Facts and Fictions about Polymorphism”[99], in which a script based on the CSD python API is used to compare atomic positions of structures within a library, and remove any structures displaying duplicate atomic coordinates. Applying this script to the libraries constructed via MGC would make the statistical data produced by MGC much more applicable to SFI.

A critical piece of contextual data for the MGC analysis of ROY is the gas-phase potential energy surface (gas-PES) for ROY which was calculation for the θ_{thio} torsion by Cruz-Cabeza et al. in 2012[86]. By plotting the θ_{thio} values onto the gas-PES of ROY, it becomes apparent that the **Y04** form adopts a high energy conformation relative to all other ROY forms, and sits close to the maximum of the PES (see Figure 45.). MGC labels the θ_{thio} torsion in **Y04** as unusual, but with

the additional context of the position this θ_{thio} value on the gas-PES of ROY, this would be enough information to label the **Y04** form as being potentially metastable. MGC alone is unable to confidently provide an evaluation of the stability of a given form, but provided the additional context of the gas-PES, a more confident conclusion as to the risk associated with the **Y04** form can be made.

4.1.2 Full Interaction Maps and Intermolecular Contacts

FIMs are heatmaps constructed using the IsoStar interaction libraries which are used to highlight areas around an input structure in which intermolecular interactions are likely to be found, what kind of interactions and their statistical likelihood. By calculating carefully selected FIMs hotspots for an input structure, and then expanding intermolecular contacts observed within the crystal lattice of an input structure, it is possible to evaluate the overall favourability of the intermolecular environment present within the input structure. When multiple polymorphs are available, this analysis can be used to rationalise differences in polymorphic stability, through comparison of the types of interactions formed, and their proximity to relevant FIMs hotspots.

Analysis of FIMs and intermolecular contacts of stable and metastable ROY polymorphs in chapter 2 revealed clear differences between the intermolecular environments of stable for forms, such as **ON** (QAXMEH), **Y** (QAXMEH01) and **O** (QAXMEH03), and the intermolecular environments of metastable forms, such as **Y04** (QAXMEH53), **R18** (QAXMEH57) and **Y19** (QAXMEH60). Stable forms tend to form contacts between favourable atoms (i.e. weak H-bond donor with weak H-bond acceptor etc.), and the position of these contacts is within or in close proximity to the corresponding FIMs hotspot. However, the ability to make these judgments for ROY polymorph is dependent on the large amount of crystallographic and polymorphic stability data available for ROY. Without these contextual data, the qualitative nature of the assessment possible via analysis of FIMs and short contacts makes it difficult to make and evaluation of the stability of a given form with any confidence.

Favourability of interacting partners and position of contact relative to FIMs hotspot are both left up to the interpretation of the user, making it difficult to discern between structures displaying similar intermolecular features (see ROY forms **Y** (QAXMEH01) and **O** (QAXMEH03) for examples). The qualitative nature of the analysis in this state leaves lots of room for potential human error, and requires strong knowledge of the intermolecular environment of the input structure(s) to be effective. A means of scoring the intermolecular environment of a given form based on the short contacts formed and their proximity to relevant

FIMs hotspots could provide a more quantitative assessment of the intermolecular environment, in a manner analogous to molecular docking.

Coordination scores can be calculated for intermolecular contacts, which would be a useful means to attach a value to the favourability of given partners involved in a contact, however, it is integration of the FIMs hotspots which holds the true potential for analysis of intermolecular environment. A means to combine coordination scores of a given contact with a score grading the proximity to/overlap with FIMs hotspots could provide a numerical evaluation of the intermolecular environment of an input structure. This would not only allow for direct comparison of the favourability of contacts observed between polymorphic forms, but also allow for more robust statistical analyses to be performed on these interactions and hence build a more complete picture of the intermolecular environment.

4.1.3 Hydrogen-Bond Propensity Modelling

HBP modelling uses the CSD to construct a regression model to analyse the H-bond donor and acceptor groups within an input structure and provide a score ranking the favourability of H-bonding networks possible using the input structure. This analysis allows for the rapid evaluation of which H-bonds are more likely to form within a crystal lattice, along with which network of H-bonds is most likely to be observed/most favourable.

Overall HBP modelling was effective at highlighting that intermolecular H-bonds are unlikely to form for the ROY structure, and that of those observed, H-bonds involving the nitrile group are more favourable than those involving the nitro group for ROY and ROY derivatives.

The only real point of contention is highlighted by **11 CB** in 3.4.2. **11 CB** is found to display three intermolecular H-bonds, and the absence of any other forms of **11 CB** suggests it could be the most stable polymorph of derivative **11**. Yet the HBP model score the observed H-bonding network in **11 CB** to be amongst the lowest calculated by the model. Upon analysis of the H-bonds in **11 CB** it is revealed that **11 CB** features a bifurcated acceptor group (the nitrile nitrogen atom), a term which the model is biased against. Further experimentation around 11 CB, or other polymorphic systems featuring bifurcated H-bond donor/acceptor groups, to explore this bias against bifurcation.

4.1.4 SFI Conclusions

It is unlikely that SFI analysis alone, in its current state, is able to confidently de-risk forms. However, the large amounts of experimental, computational and

crystallographic data available for ROY, highlight that tools in the SFI workflow are capable of identifying higher risk forms when additional contextual data are available. When looking to de-risk a form through understanding of its polymorphic landscape, the gold standards are still crystal structure prediction (CSP) and high-throughput screening (HTPS). However, both of these methods are extremely resource intensive, which inherently limits the scope of who can perform these analyses. SFI could represent an accessible middle ground between the two. When supplemented with additional computational data (i.e. gas-PES for assistance with MGC analysis), experimental (i.e. polymorphic stability data) and crystallographic data (such as the polymorph screen performed in chapter 3), SFI can be used to build a picture of the polymorphic landscape, and identify features of stable and metastable forms.

Some work could be done to the tools within the SFI workflow itself to help in its application to evaluating solid forms. A means to remove duplicate data from libraries used in MGC analysis would improve the reliability of the statistics produced. Analysis of FIMs and intermolecular short contacts shows a lot of potential in evaluating the favourability of intermolecular environments present within crystals. However, the qualitative nature of the analysis in its current state leaves a lot of ambiguity in the assessment. A more quantitative model of the intermolecular environment would allow for more confident evaluations.

Due to time constraints, a number of experiments were left out of this project, but would help provide a more thorough evaluation of the abilities of the SFI workflow to de-risk forms. Generating full crystallographic data for the polymorphs of ROY derivative synthesised in chapter three would allow for SFI analysis of these forms. Supplementing these crystallographic data with contextual data, such as gas-PES of the ROY derivatives, and polymorphic stability data of the forms of the ROY derivatives, would allow for a thorough SFI analysis of the ROY derivatives. This would also allow for a direct comparison to the SFI analysis of ROY, which is an ideal case for which relevant data is abundant, and see how SFI performs with a much smaller dataset.

SFI is an interesting toll, which, when supplemented by contextual computational and experimental data, has the potential to de-risk solid forms. This could be a useful resource for small to medium sized institutions, who lack the resources to perform CSP or HTPS studies. However, it is unlikely that SFI could replace CSP and/or HTPS any time soon.

4.2 Evaluation of the ROY Polymorphophore

A key goal for this study was to explore the polymorphic landscape of ROY using the informatics tools available within Mercury with a view to establishing a Structure-Polymorphism for the ROY polymorphophore. In this section a summary is provided of the trends in intra-, inter-, and supramolecular environment observed across the 12 ROY polymorphs within the CSD using the SFI workflow, in order to explore the SPR of the ROY polymorphophore.

4.2.1 Intramolecular Environment

The intramolecular environments of the 12 ROY polymorphs within the CSD were explored via the use of Mogul Geometry Check (MGC) in Mercury. MGC is able to rapidly produce data/measurements for all relevant intramolecular features present within the input ROY structure (bond lengths, bond angles, intramolecular torsions and ring geometries), and provides a library of structures containing similar structural features for statistical comparison. For the ROY polymorphs, the only intramolecular feature found to vary significantly across the 12 ROY forms was the torsion between the two aromatic rings about the bridging amine, referred to as θ_{thio}.

Using MGC, a library of ROY-like structures was constructed to explore the distribution of θ_{thio} across the CSD. Using this information, analysis of the ROY-like structures within the library highlights a number of other potentially metastable forms. This could suggest that, for ROY, planar formations, or outside of the θ_{thio} ranges, could be unfavourable and highlight metastable forms.

These observations were largely found to hold true for the ROY derivatives synthesised in chapter 3. The distribution of θ_{thio} values observed within the crystal lattices of the forms of derivatives **1** – **12** is largely similar to that observed for ROY-like structures in chapter 2, with the exception of **9 DR**, **10 RP**, **12 RN** and **12 ON**, all of which are forms of derivatives in which the aminothiophene ring of the ROY structure is modified. These forms are unusually planar for ROY ROY-like structures, and more planar than any ROY polymorph in the CSD, with θ_{thio} values < 2º.

For **12 RN** and **12 ON**, this is explained by the intramolecular H-bond observed between the bridging amine group and the newly introduced aldehyde, but the planarity of **9 DR** and **10 RP** could suggest a link between aminothiophene modification and lower θ_{thio} values for ROY derivatives.

4.2.2 Intermolecular Environment

A combination of Full Interaction Maps (FIMs) and visualisation of short contacts has been used to compare the intermolecular environment of three stable ROY

forms (**ON** (QAXMEH), **Y** (QAXMEH01) and **O** (QAXMEH03)) with three metastable ROY forms (**Y04** (QAXMEH53), **R18** (QAXMEH57) and **Y19** (QAXMEH60)) in order to evaluate factors which influence the intermolecular environment of ROY polymorphs. Broadly, it was found that a small number of intermolecular contacts located within/in close proximity to the dominant FIMs hotspots is an indicator of stable ROY polymorphs, and vice versa.

Specifically, it is observed that the nitrile environment of ROY polymorphs is the most influential intermolecular factor on polymorphic stability, followed by the environment of the nitro group. This is also highlighted by the relative size and darkness/density of the FIMs hotspots associated with the nitrile and nitro groups relative to those produced for the remainder of the ROY scaffold.

The lack of polymorph data for all derivatives except for derivative **12** limits the applicability of intermolecular analysis for the ROY derivatives synthesised in Chapter 3, however, the trends in influence of the nitrile and nitro groups seem to largely hold true, the only exception being **11 CB**, in which the nitro group present in ROY is converted to a primary amine. To date, no additional forms of **11 CB** have been reported since its discovery in 2010[95], and **11 CB** features a nitrile group involved in two intermolecular H-bonds. This results in a stable form of a ROY derivative which features favourable primary and secondary amine environment in **11 CB**, and a less favourable nitrile environment.

4.2.3 Supramolecular Environment

In section 2.3 Hydrogen-Bond Propensity (HBP) modelling was used to evaluate the supramolecular environments of the 12 ROY polymorphs within the CSD and the 12 ROY derivatives synthesised in Chapter 3. These analyses predict that ROY polymorphs and most of the ROY derivatives are unlikely to form many intermolecular H-bonds, if any at all, as highlighted by the low propensity and coordination scores for the H-bond donor and acceptor groups, along with the low number of intermolecular H-bonds observed within their respective crystal lattices. This s likely due to the influence of the intramolecular H-bond between the nitro group and bridging amine which is present in each of the ROY polymorphs and derivatives discussed, except for **11 CB**, which exhibits different behaviour.

Despite this, analysis of the thermal data of the 12 ROY polymorphs within the CSD revealed that the presence of an intermolecular H-bond between the nitrile and bridging amine groups could be indicative of stable forms (**Y** (QAXMEH01), **O** (QAXMEH03) and **YT04** (QAXMEH12)), whilst an intermolecular H-bond between the nitro and bridging amine groups was associated with the metastable

YN (QAXMEH04) form. Zero intermolecular H-bonds was found to have no strong correlations with stable or metastable polymorphs.

Of the 12 ROY polymorphs and 12 ROY derivatives discussed in this work, 11 CB is the only scaffold which displays a different number of H-bond donor and acceptor groups; displaying three donors (one from primary amine, two from secondary amine) and two acceptors (nitrile nitrogen and primary amine nitrogen), compared to the one donor (primary amine) and three acceptors displayed by the remainder of the structures discussed. **11 CB** features three intermolecular H-bonds which form a sort of chain, a much more complex H-bonding network than observed (or even possible) in other structures discussed in this work

4.2.4 Polymorph Screen

In 3.4 a polymorph screen was designed and applied to crystals of the 12 ROY derivatives synthesised in Chapter 3. The goal of this screen was to verify the ability of the ROY scaffold to produce structural derivatives which display polymorphism. Similar to the studies performed on the fenamate scaffold by Lopez-Mejias et al. in 2015[94] and NPBA by Bhandary et al. in 2019[100], this screen was designed to confirm the ROY scaffold as a polymorphophore.

ROY and it polymorphs are observed to exhibit colour polymorphism, and during synthesis of the 12 ROY derivatives, it was found that similar colour polymorphism was also observed for a number of ROY derivatives. Given the restrictions and time and resources, observation of forms of different colour and/or habit was used as an initial detection method for the identification of new forms during this polymorph screen. However, upon performing the screening experiments it was found that only derivatives **1** – **8** were observed to undergo significant colour polymorphism. Derivative **12** was found to crystallise in two needle forms with subtle differences in colour, whilst no significant colour change was observed for derivatives **9** – **11**.

PXRD patterns obtained from samples generated during the polymorph screen were compared to simulated PXRD patterns from the crystals generated and characterised in Chapter 3 in order to confirm the presence of new polymorphic forms. However, this method produced a number of issues. Usually, a threshold of < 95% pattern match is use to confirm the presence of a new polymorphic form. Direct comparison of experimental patterns of crystals generated in Chapter 3 using the setup used to characterise the experiments from the polymorph screen with simulated PXRD patterns from the same crystal form yielded a maximum match of 93%. There may be a number of reasons for thus discrepancy, including low crystallinity of crystals produced from the polymorph screen, or too little

material used per crystallisation experiment Furthermore, comparison of simulated PXRD patterns to experimentally obtained patterns may also explain the low percentage match. This made identification and confirmation of new polymorphic forms more difficult, particularly for derivatives **9** – **11**.

Despite this, the polymorph screen was able to identify and confirm new polymorphic forms of derivatives **1** – **8**, with potential polymorphs of derivatives **9** and **10** identified but not verified. The screen was also able to identify both forms of derivative **12** which were identified in Chapter 3, with PXRD also suggesting another undiscovered form, but yet to be verified. Of the 12 ROY derivatives synthesised in Chapter 3, nine were confirmed to display polymorphism (75%). In addition to those nine confirmed cases of polymorphism, derivatives **9** and **10** were also found to display potential polymorphic forms, but unverified due to the lack of observable colour polymorphism. Derivative **11** (XABXEF) was the only derivative for which no significant evidence for polymorphism was found.

Of the 12 derivatives studied, nine were confirmed to display polymorphism, and a further two derivatives show some evidence of new polymorphic forms. These data would suggest that the ROY scaffold has a high probability of producing structural derivatives which also display polymorphism.

4.2.5 Conclusions and Future Perspectives

Based on the data generated in Chapter 3, previous studies into ROY derivatives, and the criteria laid out by Lopes-Mijas et al. and Bhandary et al., it is highly likely the ROY scaffold is a polymorphophore. This takes the total number of confirmed polymorphophores confirmed through screening to be three: fenamates, NPBA and ROY. However, due to a lack of data surrounding polymorphophores, the implications of this classification are still unclear.

With three confirmed cases of polymorphophores identified at current date, a next logical step would be to perform comparative studies between ROY, the fenamate scaffold and NPBA, in order to establish a set of features commonly held by polymorphophores. All three of the compounds listed contain two π-systems connected by a bridging group, whilst studies into all three scaffolds suggest that introduction of small groups around the benzene ring(s) present in the scaffold are more likely to produce derivatives displaying polymorphism. Understanding the interplay between the p-systems and the influence of these substitutions may hold the key to understanding the origins of polymorphophores.

Given the structural similarities between the ROY, fenamate and NPBA scaffolds, a list of commonly held traits of polymorphophores is beginning to form.

Expansion of this list of traits could be used in an attempt to designed new polymorphophore structures from first principles.

Finally, despite confirmation of the ROY scaffold as a polymorphophore, we are no closer to understanding the nature of polymorphophores outside of their unusual polymorphic properties. Polymorphism is a generally unwanted trait discovered at later stages of the development of chemical products. Further studies into polymorphophores may be useful in understanding factors present in a chemical scaffold which may influence the accessibility of alternate polymorphic forms, but at this stage, more studies regarding the nature of polymorphophores are required.

Chapter 5 Experimental Methods

5.1 Solid-Form Informatics Workflow

All informatics tools used in this work can be accessed through Mercury (2023.2.0, Build 382240) which is provided for free by the Cambridge Crystallographic Data Centre (CCDC), and all analysis were performed using Cambridge Structural Database (CSD) version 5.44 (April 2023).

5.1.1 Mogul Geometry Check

Mogul Geometry Check (MGC) is accessible via the “CSD-Core” section of Mercury. In order to attain reliable statistics for the different ROY polymorphs it is important to use the “kekulised” representation of the thiophene ring (see Figure 39).

When performing MGC analysis, the following settings were applied:

Fragment types = “Bond Length”, “Valence Angle”, “Torsion Angle”.

No filters applied.

Find similar fragments if number of exact matches is less than – Bonds = 15, Angles = 15, Torsions = 40.

5.1.2 Full-Interaction Maps and Intermolecular Contacts

Full-Interaction Maps (FIMs) are used to explore the intermolecular environments of the 12 ROY polymorphs and 12 ROY analogues studied in this work. FIMs are accessed via the “CSD-Materials” tab in Mercury by selecting “Full Interaction Maps”. This produces pop-out window in which settings for the FIMs can be altered, such as the probe species (and their associated colours) used to produce FIMs contour plots. In this work the uncharged NH nitrogen (blue), carbonyl oxygen (red) and aromatic CH carbon (orange/brown) probes were used to produce FIMs for all ROY polymorphs and analogues. Default settings were used for all FIMs analyses performed during this work.

FIMs are calculated for the species loaded in the main window of Mercury .Once calculated FIMs hotspots can be seen surrounding the molecule present in the main window of Mercury. The colour and shade of the contours correspond to the probe found at a given distance from the central molecule and the statistical likelihood of finding the probe at that position.

In order to investigate the intermolecular environment of the central molecule, the H-bonds and intermolecular short contacts must be explored. To visualise these

contacts, navigate to the “Display” toolbar located at the bottom of the main window of Mercury. Ticking the “Short Contact” and “H-Bond” boxes produces a set of red and blue dashed lines which indicate the presence of intermolecular short contacts and H-bonds respectively. In this work, default settings for short contact and H-bond definitions were used for all FIMs analyses.

Red lines delineating short contacts can be found originating from all atoms involved in intermolecular short contacts/interactions, and terminate with a red “X” highlighting where the short contact is found. Clicking on these red lines expands the contacts to show the position of the other molecule involved in the intermolecular interaction, with the red “X” being replaced by the other atom involved in the short contact. Analysis of the types of intermolecular interactions observed, and their proximity to the corresponding FIMs hotspot provides information to evaluate the quality of intermolecular interactions formed in packing of the species in question.

5.1.3 Hydrogen-Bond Propensity Modelling

Hydrogen-Bond Propensity (HBP) modelling is a tool used to evaluate the supramolecular environments found within a given crystal lattice. The HBP modelling tool can be accessed via the “Polymorph Assessment” section located within the “CSD-Materials” tab in Mercury. Clicking the HBP modelling tool produces the “Propensity Prediction Wizard” pop-out window, where users can configure settings relating to the construction of the HBP model. In this work the default functional group library and default Hydrogen bond definition was used for all HBP models. All models were constructed using the CSD Version 5.44, Jun23.

At the bottom right of the “Propensity Prediction Wizard” window a representation of the molecule to be subject to HBP modelling can be found, with all functional groups highlighted. When performing HBP analysis for ROY polymorphs and ROY analogues, most consistent results were obtained using the kekulised representation of the thiophene ring (see Figure 39 for further information).Once HBP settings are in place, click “Next” to continue and open the “Generate Fitting Data” window.

The next step is to select structures from the CSD to act as fitting data for the HBP model. In this work, data generation was truncated at 5000 items for each HBP model. Once ready, click “Generate” to generate fitting data. This produces a table in the bottom left of the “Generate Fitting Data” window listing the functional groups present in the input structure, along with a count of how many of each functional group are present within the fitting dataset. To the right of this table is a column labelled “Advice”, which evaluates the number of functional

groups present within the fitting data and suggests whether the number is sufficient for reliable statistics. Typically, around 300 entries is sufficient.

The next step is to analyse the structures contained within the fitting set.by clicking the “Analyse” button. This analysis investigates the structures within the fitting set to see whether the functional groups present table behave as H-bond donors or acceptors, and produces a table highlighting these statistics. For the ROY scaffold, the HBP analysis highlights one donor group (atom_0_of_sec_amine_1) and four acceptor groups (atom_0_of_sec_amine_1, atom_1_of_cyclic_thioether, atom_2(3)_of_ar_nitro and atom_2_of_cyano). However, due to a lack of instances of atom_0_of_sec_amine_1 acting as an acceptor group, this feature was left out of the HBP model of ROY analogues

5.2 Synthetic Methods

Reagents and solvents were obtained from commercial suppliers and used without further purification, unless otherwise stated. Solvents used were HPLC or analytical grade. Thin layer chromatography was performed on aluminium backed silica gel supplied by Merck, visualised using an ultraviolet lamp. Flash column chromatography was performed using silica gel 60 (40 – 63 um particles). Hydrogen and carbon NMR data were collected on a Bruker Advance III 500. All shifts recorded against an internal standard against an internal standard of tetramethyl silane. Chloroform-*d* used for NMR solvent was obtained from Sigma-Aldrich. ^{1}H NMR data is reported in the following format: ppm (splitting pattern, coupling constant (Hz), number of protons, proton assignment). Signal assignments were deduced with aid of MestReNova and DEPT 135. LC-MS (liquid chromatography-mass spectrometry) data were recorded on a Donex Ultimate 3000 LC system with a MeCN/H_2O with + 0.1% formic acid gradient. HRMS (High resolution mass spectrometry) data were recorded on a Bruker MaXis impact spectrometer using electron spray ionisation. Infrared spectra were recorded on a Perkin-elmer one FTIR spectrometer. Melting points were recorded using a (ask Adrian) Differential Scanning Calorimeter. LC-MS data are not reported for compounds for which HPLC data have been provided.

5.2.1 General Experimental Procedures

5.2.1.1 Method A: Palladium on Activated Charcoal Hydrogenation

The procedure outlined below is modified from a literature source[101]. A two-neck round-bottomed flask was fitted with a magnetic stirrer bar, along with 5-methyl-

2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile (6.0 mmol) and palladium on activated charcoal (0.6 mmol). The vessel was sealed using suba seals and dichloromethane (50 mL) was added through the side neck of the flask. A nitrogen line was attached and an outlet needle attached to the side neck of the flask. The vessel was allowed to purge with nitrogen gas for 20 minutes whilst stirring was applied to allow 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile to dissolve.

After purging for 20 minutes, the nitrogen line was disconnected and a hydrogen balloon was attached to the vessel via the top neck of the flask. The valve was opened, allowing for the flow of hydrogen into the vessel. An outlet needle was attached to the side neck of the flask to allow the vessel to saturate with hydrogen gas for 20 minutes. After 20 minutes, the outlet needle was removed and the reaction was left to stir at RT. The reaction was monitored by sampling the reaction mixture using a syringe down the side neck of the vessel. Reaction mixture was passed through a pipette column filled with celite to remove and Pd/C, and the resulting solution was analysed using thin-layer chromatography (TLC) using a 4;1 mixture of ethyl acetate and hexane.

Once TLC showed no signs of starting material, the hydrogen balloon was disconnected and the nitrogen line was reconnected along with an outlet needle. The vessel was allowed to be purged for an hour to ensure total removal of hydrogen gas from the vessel. After an hour, the nitrogen line was removed along with the suba seals from the vessel. The reaction mixture was then filtered through celite (ca. 100 cm^3) under reduced pressure, and the celite was washed with 4 x 50 mL dichloromethane. The resulting solution was then concentrated under reduced pressure to yield crude 11. Pure 11 was isolated via flash column chromatography using a 4;1 ethyl acetate/hexane column solvent.

5.2.1.2 Method B: Diisobutylaluminium Hydride Reduction

A flask is dried, flushed with nitrogen and filled with 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile (4 mmol) which is dissolved in dry DCM (50 mL) with stirring. The solution was cooled to -110^{o}C by submerging the reaction vessel in an acetone/dry ice bath and left to climatise for 30 minutes. 1M DIBAL solution (1 eq) was added dropwise over the course of 20 minutes in order to maintain the temperature inside the vessel. The reaction was left to stir for an

hour under cryogenic conditions before the acetone/dry ice bath was removed, and the reaction vessel was allowed to warm to 0°C in an ice/water slurry. The reaction was maintained at 0°C and left to stir vigorously for 4 hours before excess DIBAL was quenched with water (1 – 5 mL) and methanol (3 – 15 mL). 10% Rochelle's salt solution (5 – 30 mL) was added along with water (100 mL) and EtOAc (150 mL). The aqueous layer was washed with EtOAc (3 x 100 mL), before the organic layers were combined, washed with water (3 x 100 mL), brine (50 mL), dried ($MgSO_4$) and concentrated under reduced pressure to yield crude product.

Pure product was obtained via flash column chromatography using a 4:1 EtOAc/hexane column solvent.

5.2.1.3 Method C: Gewald Aminothiophene Synthesis

The procedure outlined below is modified from a literature source[102]. In a dry round-bottomed flask malononitrile (12 mmol, 793 mg), ground octasulfur (12.5 mmol, 3200 mg) and L-proline (1.2 mmol, 140 mg) were added along with a magnetic stirrer bar. The flask was sealed and purged with nitrogen for 30 minutes. Dry DMF (130 mL) was added and the mixture was left to stir under nitrogen for 30 minutes. Selected carbonyl reagent (acetone = 12 mmol, 760 μL/butanone = 12 mmol, 1080 μL) was then added to the solution dropwise. The reaction was then left to stir for 6 hours, or until malononitrile was no longer present under TLC.

Upon completion the reaction mixture was diluted into 800 mL EtOAc. DFM was removed by washing the diluted mixture with 1M LiCl (aq) (8 x 100 mL). The organic layer was then washed with water (3 x 300 mL), brine (150 mL) and dried ($MgSO_4$), before being concentrated under reduced pressure. Crude product was isolated via dissolving in toluene and recrystallised with hexane at 0°C overnight.

Pure product was obtained via flash column chromatography using 100:1 DCM/methanol as column solvent.

5.2.1.4 Method D: S_NAr Synthesis

In a dry two-neck round-bottomed flask fitted with an addition funnel KOH pellets (9 mmol, 505 mg) and a magnetic stirrer bar are added. The flask was sealed and purged with nitrogen for 30 minutes before dry DMF (30 mL) was added. In

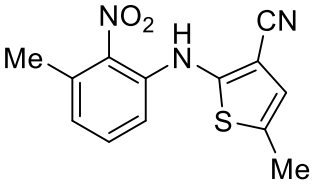
a separate round-bottomed flask fitted with a magnetic stirrer bar aminothiophene reagent (6 mmol) is added, before the flask is sealed and purged with nitrogen for 30 minutes. Dry DMF (30 mL) is added to dissolve the aminothiophene reagent, before 1,2-fluoronitrobenzene (6.2 mmol, 650 μL) was added dropwise. This solution was then added to the addition funnel, and added dropwise to the KOH solution over 30 minutes. The reaction was left to stir under nitrogen for 24 hours.

After 24 hours, the reaction mixture was diluted into EtOAc (300 mL) and DMF removed by washing with 1M LiCl (aq) (8 x 50 mL). The organic layer was then washed with water (3 x 100 mL), brine (50 mL) and dried ($MgSO_4$), before being concentrated under reduced pressure to yield crude product.

Pure product was isolated via flash column chromatography. See individual compounds for detailed purification methods.

5.2.2 Synthetic Procedures

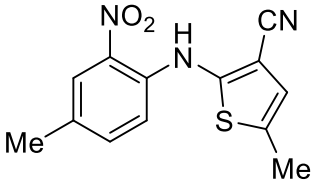
5-methyl-2-[(3-methyl-2-nitrophenyl)]amino-3-thiophenecarbonitrile (**1**) [SLM143]



Synthesised *via* method **D** from 5-methyl-2-amino-3-thiophenecarbonitrile (12.0 mmol, 1670 mg) and 3-fluoro-2-nitrotoluene (12.0 mmol, 1910 mg). Compound **1** obtained as orange needles (1310 mg, 40 %).

MP; 111.5 – 112.5 °C. **^{1}H NMR (500 MHz, chloroform-*d*)**; δ 2.35 (d, *J* = 1.3 Hz, 3H, C5-C**H**$_3$), 2.44 (s, 3H, C3'-C**H**$_3$), 6.61 (q, *J* = 1.3 Hz, 1H, C4-**H**), 6.81 (d, *J* = 8.0 Hz, 1H, C6'-**H**), 7.06 (d, *J* = 8.0 Hz, 1H, C4'-**H**), 7.24 (t, *J* = 8.0 Hz, 1H, C5'-**H**), 8.07 (s, 1H, N-**H**).**^{13}C NMR (125 MHz, chloroform-*d*)**; δ 15.5 (C5- CH_3), 20.5 (C3'-CH_3), 101.2 (C5), 114.0 (C4), 115.1 (C3'), 122.1 (C6'), 123.5 (C4'), 124.7 (C5'), 132.9 (C2), 133.4 (C1'), 135.2 (C3), 138.3 (C2'), 151.4 (C≡N). **ν_{max} / cm^{-1}**; (C≡N), (C=C), (N-H), (N=O). **HRMS (ESI)**; calculated for $C_{13}H_{12}N_3O_2S^+$ $[M+H]^+$ 274.0645, found 274.0644. **HPLC**; RT = 3.282 - 3.384 min, area = 100%.

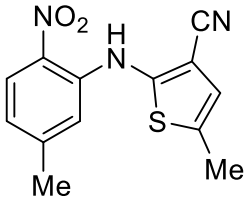
5-methyl-2-[(4-methyl-2-nitrophenyl)]amino-3-thiophenecarbonitrile (**2**) [SLM166]



Synthesised *via* method **D** from 5-methyl-2-amino-3-thiophenecarbonitrile (6.0 mmol, 835 mg) and 4-fluoro-3-nitrotoluene (6.0 mmol, 939 mg). Compound **2** obtained as red needles (982 mg, 60 %).

MP; 114.5 - 115.0 °C. **^{1}H NMR (500 MHz, chloroform-*d*)**; δ 2.35 (d, *J* = 0.7 Hz, 3H, C4'-C**H**$_3$), 2.46 (d, *J* = 1.2 Hz, 3H, C4-C**H**$_3$), 6.75 (q, *J* = 1.2 Hz, 1H, C4-**H**), 7.14 (d, *J* = 8.0 Hz, 1H, C6'-**H**), 7.34 (d, *J* = 8.0 Hz, 1H, C5'-**H**), 8.05 (q, *J* =0.7 Hz, 1H, C3'-**H**), 9.54 (s, 1H, N-**H**).**^{13}C NMR (125 MHz, chloroform-*d*)**; δ 15.6 (C5- CH_3), 20.5 (C4'-CH_3), 103.6 (C5), 113.9 (C4), 116.2 (C4'), 122.1 (C6'), 123.8 (C3'), 126.2 (C5'), 130.1 (C2), 135.3 (C1'), 137.4 (C3), 138.9 (C2'), 160.6 (C≡N). **ν_{max} *I* cm^{-1}**; (C≡N), (C=C), (N-H), (N=O). **HRMS (ESI)**; calculated for $C_{13}H_{12}N_3O_2S^+$ $[M+H]^+$ 274.0645, found 274.0644. **HPLC**; RT = 3.282 - 3.384 min, area = 99.4 %.

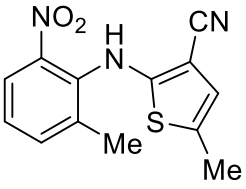
5-methyl-2-[(5-methyl-2-nitrophenyl)]amino-3-thiophenecarbonitrile (**3**) [SLM168]



Synthesised *via* method **D** from 5-methyl-2-amino-3-thiophenecarbonitrile (6.0 mmol, 833 mg) and 3-fluoro-4-nitrotoluene (6 mmol, 991 mg). Compound **3** obtained as red prisms (1060 mg, 64 %).

MP; 144.8 – 145.5 °C. **^{1}H NMR (500 MHz, chloroform-*d*)**; δ 2.36 (d, *J* = 0.7 Hz, 3H, C5'-C**H**$_3$), 2.48 (d, *J* = 1.2 Hz, 3H, C5-C**H**$_3$), 6.78 (q, *J* = 1.2 Hz, 1H, C4-**H**), 6.96 (d, *J* = 8.0 Hz, 1H, C6'-**H**), 7.26 (d, *J* = 8.0 Hz, 1H, C5'-**H**), 8.14 (q, *J* = 0.7 Hz, 1H, C3'-**H**), 9.66 (s, 1H, N-**H**).**^{13}C NMR (125 MHz, chloroform-*d*)**; δ 15.6 (C5-CH_3), 20.5 (C4'-CH_3), 103.6 (C5), 113.9 (C4), 116.2 (C4'), 122.1 (C6'), 123.8 (C3'), 126.2 (C5'), 130.1 (C2), 135.3 (C1'), 137.4 (C3), 138.9 (C2'), 160.6 (C≡N). **ν_{max} *I* cm^{-1}**; (C≡N), (C=C), (N-H), (N=O). **HRMS (ESI)**; calculated for $C_{13}H_{12}N_3O_2S^+$ $[M+H]^+$ 274.0645, found 274.0643. **HPLC**; RT = 3.46 - 3.58 min, area = 98.8 %.

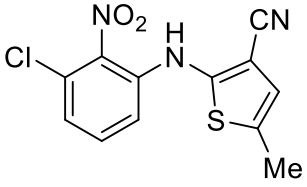
5-methyl-2-[(6-methyl-2-nitrophenyl)]amino-3-thiophenecarbonitrile (**4**) [SLM142]



Synthesised *via* method **D** from 5-methyl-2-amino-3-thiophenecarbonitrile (12.0 mmol, 1690 mg) and 2-fluoro-3-nitrotoluene (6.0 mmol, 1910 mg). Compound **4** obtained as red prisms (1420 mg, 44 %).

MP; 129.6 – 131.5 °C. **^{1}H NMR (500 MHz, chloroform-*d*)**; δ 2.28 (d, *J* = 0.8 Hz, 3H, C6'-C**H**$_3$), 2.39 (d, *J* = 1.5 Hz, 3H, C5-C**H**$_3$), 6.68 (q, *J* = 1.5 Hz, 1H, C4-**H**), 7.79 (m, 2H, C3'-**H** + C4'-**H**), 7.98 (d, *J* = 0.8 Hz, 1H, C5'-**H**), 9.47 (s, 1H, N-**H**).**^{13}C NMR (125 MHz, chloroform-*d*)**; δ 15.4 (C5-CH_3), 19.1 (C6'-CH_3), 97.1 (C5), 114.2 (C4), 122.1 (C6'), 123.4 (C4'), 123.9 (C3'), 124.4 (C5'), 131.0 (C2), 134.7 (C1'), 136.8 (C3), 137.9 (C2'), 156.6 (C≡N). **HRMS (ESI)**; calculated for $C_{13}H_{12}N_3O_2S^+$ $[M+H]^+$ 274.0645, found 274.0643. **HPLC**; RT = 3.181 - 3.297 min, area = 99.7 %.

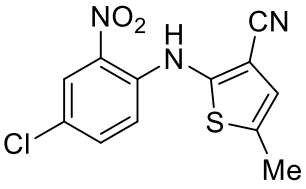
5-methyl-2-[(3-chloro-2-nitrophenyl)]amino-3-thiophenecarbonitrile (**5**) [SLM96]



Synthesised *via* method **D** from 5-methyl-2-amino-3-thiophenecarbonitrile (6.0 mmol, 831 mg) and 2-chloro-6-fluoronitrobenzene (6.0 mmol, 1070 mg). Compound **5** obtained as orange needles (950 mg, 54 %).

MP; 130.7 – 132.9 °C. **^{1}H NMR (500 MHz, chloroform-*d*)**; δ 2.37 (d, *J* = 1.4 Hz, 3H, C5-C**H**$_3$), 6.64 (q, *J* = 1.4 Hz, 1H, C4-**H**), 7.02 (m, 2H, C4'-**H** + C6'-**H**), 7.25 (dd, *J* = 10.4, 0.7 Hz, C5'-**H**), 7.48 (s, 1H, N-**H**). **^{13}C NMR (125 MHz, chloroform-*d*)**; 15.2 (C5-CH_3), 102.9 (C5), 113.5 (C4), 114.4 (C3'), 115.5 (C5'), 123.5 (C6'), 128.7 (C2), 132.7 (C1'), 135.0 (C3), 139.2 (C2'), 150.0 (C≡N).. **HRMS (ESI)**; calculated for $C_{13}H_9N_3Cl^{35}O_2S^+$ $[M+H]^+$ 294.0095, found 294.0099. **HPLC** RT = 3.35 – 3.45mins, Area = 95.4%;

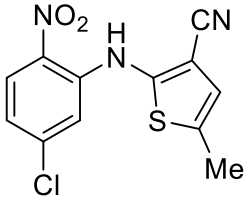
5-methyl-2-[(4-chloro-2-nitrophenyl)]amino-3-thiophenecarbonitrile (**6**) [SLM166]



Synthesised *via* method **D** from 5-methyl-2-amino-3-thiophenecarbonitrile (6.0 mmol, 835 mg) and 2-chloro-5-fluoronitrobenzene (6.0 mmol, 1070 mg). Compound **6** obtained as red prisms (1110 mg, 63 %).

MP; 141.4 – 142.3 °C.. **^{1}H NMR (500 MHz, chloroform-*d*)**; δ 2.42 (d, *J* = 1.3 Hz, 3H, C5-C**H**$_3$), 6.73 (q, *J* = 1.3 Hz, 1H), 7.05 (d, *J* = 9.0 Hz, 1H, C6'-**H**), 7.39 (dd, *J* = 9.0 + 2.4 Hz, C5'-**H**), 8.18 (d, *J* = 2.4 Hz, 1H, C3'-**H**), 9.42 (s, 1H, N-**H**). **^{13}C NMR (125 MHz, chloroform-*d*)**; 15.7 (C5-CH_3), 105.6 (C5), 113.5 (C4), 117.5 (C4'), 124.2 (C6'), 124.8 (C4), 126.0 (C3'), 134.0 (2), 136.3 (C1'), 137.1 (C3), 140.1 (C2'), 148.1 (C≡N). **HRMS (ESI)**; calculated for $C_{13}H_9N_3Cl^{35}O_2S^+$ $[M+H]^+$ 294.0095, found 294.0098. **HPLC** RT = 4.03 – 4.16 mins, Area = 99.8%;

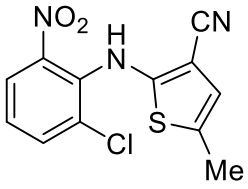
5-methyl-2-[(5-chloro-2-nitrophenyl)amino]-3-thiophenecarbonitrile (**7**) [SLM94]



Synthesised *via* method **D** from 5-methyl-2-amino-3-thiophenecarbonitrile (12.0 mmol, 1860 mg) and 2-chloro-4-fluoronitrobenzene (12.0 mmol, 2180 mg). Compound **7** obtained as yellow prisms (891 mg, 25 %).

MP; 152.1 – 153.0 °C.. **^{1}H NMR (500 MHz, chloroform-*d*)**; δ 2.44 (d, *J* = 1.2 Hz, 3H, C5-C**H**$_3$), 6.76 (q, *J* = 1.2 Hz, 1H, C4-**H**), 6.85 (dd, *J* = 9.1 + 2.1 Hz, 1H, C4'-**H**), 7.02 (d, *J* = 2.1 Hz, 1H, C6'-**H**), 8.13 (d, *J* = 9.1 Hz, C3'-**H**), 9.42 (s, 1H, N-**H**). **^{13}C NMR (125 MHz, chloroform-*d*)**; 15.7 (C5-CH_3), 106.2 (C5), 113.3 (C4), 115.6 (C5'), 120.2 (C4'), 124.3 (C6'), 128.1 (C3'), 133.2 (C2), 137.7 (C1'), 142.5 (C3), 143.0 (C2'), 147.4 (C≡N). **HRMS**; m/z (ES) found $M+Na^+$ 315.9918; $C_{12}H_8ClN_3NaO_2S$ requires $M+Na^+$ 315.9918. **HPLC** RT = 3.56 – 3.69 mins, Area = 99.1 %.

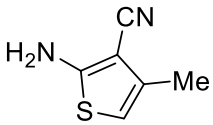
5-methyl-2-[(6-chloro-2-nitrophenyl)]amino-3-thiophenecarbonitrile (**8**) [SLM168]



Synthesised *via* method **D** from 5-methyl-2-amino-3-thiophenecarbonitrile (6.0 mmol, 832 mg) and 2-chloro-3-fluoronitrobenzene (6.0 mmol, 1090 mg). Compound **8** obtained as red needles (786 mg, 45 %).

MP; 108.8 – 110.0 °C.. **^{1}H NMR (500 MHz, chloroform-*d*)**; δ 2.31 (d, *J* = 1.3 Hz, 3H, C5-C**H**$_3$), 6.53 (q, *J* = 1.3 Hz, 1H, C4-**H**), 7.09 (t, *J* = 8.0 Hz, 1H, C4'-**H**), 7.60 (dd, *J* = 8.0 + 1.5 Hz, 1H, C5'-**H**), 8.02 (dd, *J* = 8.0 + 1.5 Hz, C3'-**H**), 8.39 (s, 1H, N-**H**). **^{13}C NMR (125 MHz, chloroform-*d*)**; 15.5 (C5-CH_3), 100.1 (C5), 113.9 (C4), 123.4 (C6'), 123.5 (C4'), 125.0 (C5'), 128.6 (C3'), 133.6 (C2), 136.2 (C1'), 137.1 (C3), 141.6 (C2'), 153.3 (C≡N). **HRMS (ESI)**; calculated for $C13H9N3Cl^{35}O2S^{+}$ $[M+H]^{+}$ 294.0095, found 294.0098. **HPLC**; RT = 3.11 - 3.21 mins, Area = 96.5 %.

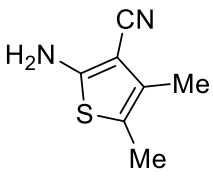
4-methyl-2-aminothiophene-3-carbonitrile (**9a**) [SLM113]



Synthesised *via* method **C** from acetone (12 mmol, 0.9 mL). Pure product was isolated via flash column chromatography using 100:1:0.1 DCM/methanol/Et_3N column solvent. Compound **9a** obtained as an off-white plate (550 mg, 25%).

^{1}H NMR (500 MHz, chloroform-*d*); δ 2.12 (d, *J* = 1.2 Hz, 3H, C4-C**H**$_3$), δ 4.57 (bs, 2H, N-**H**), δ 5.90 (q, *J* = 1.2 Hz, 1H, C5-**H**). **^{13}C NMR (125 MHz, chloroform-*d*)**; 13.1 (C4-CH_3), 91.8 (C5), 115.3 (C4), 116.9 (C2), 128.1 (C3), 156.9 (C≡N). **HPLC**; RT = 4.263 – 4.353 mins, Area = 75.3%

4,5-dimethyl-2-aminothiophene-3-carbonitrile (**10a**) [SLM114]

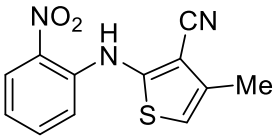


Synthesised *via* method **C** from 2-butanone (12 mmol, 1.1 mL). Pure product was isolated via flash column chromatography using 4.1 EtOAc/hexane column solvent. Compound **10a** obtained as an off-white plate (851 mg, 32 %).

^{1}H NMR (500 MHz, chloroform-*d*); δ 2.00 (q, *J* = 0.8 Hz, 3H, C5-C**H**$_3$), δ 2.09 (q, *J* = 0.8 Hz, 3H, C4-C**H**$_3$), δ 4.41 (bs, 2H, N-**H**). **^{13}C NMR (125 MHz, chloroform-**

***d*)**; 12.4 (C5-CH_3), 12.8 (C4-CH3), 91.0 (C5), 115.6 (C4), 117.2 (C2), 129.7 (C3), 158.8 (C≡N). **ν_{max} / cm^{-1}**; (C≡N), (C=C), (N-H). **HPLC**; RT = 4.339 – 4.440 mins, Area = 99.23%.

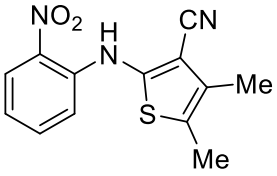
4-methyl-2-[(2-nitrophenyl)amino]thiophene-3-carbonitrile (**9**) [SLM141]



Synthesised *via* method **D** from **9a** (6.0 mmol, 821 mg) and 1,2-fluoronitrobenzene (7.0 mmol, 800 μL). Compound **9** obtained as dark red prisms (315 mg, 20 %).

MP;118.5 – 119.7 °C . **^{1}H NMR (500 MHz, chloroform-*d*)**; δ 2.29 (d, *J* = 1.3 Hz, 3H, C4-C**H**$_3$), δ 6.64 (q, *J* = 1.3 Hz, 1H, C5-**H**), δ 6.92 (m, 1H, C4'-**H**), δ 7.22 (dd, *J* = 1.3 + 8.5 Hz, 1H, C6'=**H**), δ 7.47 (m, 1H, C5'-**H**), δ 8.19 (dd, *J* = 1.6 + 8.5 Hz, 1H, C3'-**H**), δ 9.73 (s, 1H, N-**H**). **^{13}C NMR (125 MHz, chloroform-*d*)**; 21.1 (C4-CH_3), 105.8 (C4), 113.6 (C5), 115.4 (C4'), 116.2 (C5'), 120.2 (C6'), 126.7 (C3'), 134.4 (C3), 136.3 (C2), 137.2 (C2'), 140.6 (C1'), 151.2 (C≡N). **HRMS (ESI)**; calculated for $C_{12}H_9N_3O_2S^+$ $[M+H]^+$ 260.0645, found 260.0643. **HPLC** RT = 2.654 – 2.778 mins, Area = 99.76%;

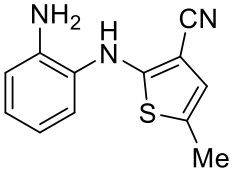
4,5-dimethyl-2-[(2-nitrophenyl)amino]thiophene-3-carbonitrile (**10**) [SLM156]



Synthesised *via* method **D** from **10a** (6.0 mmol, 940 mg) and 1,2-fluoronitrobenzene (7.0 mmol, 800 μL). Compound **10** obtained as dark red needles (725 mg, 43 %).

MP; 116.5 – 118.2 °C . **^{1}H NMR (500 MHz, chloroform-*d*)**; δ 2.17 (s, 3H, C5-C**H**$_3$), 2.28 (s, 3H, C4-C**H**$_3$), 6.88 (m, 1H, C4'-**H**), 7.14 (dd, *J* = 1.3 + 8.6 Hz, 1H, C6'-**H**), 7.44 (m, 1H, C5'-**H**), 8.17 (dd, *J* = 1.6 + 8.6 Hz, 1H, C3'-**H**), 9.55 (s, 1H, N-**H**). **^{13}C NMR (125 MHz, chloroform-*d*)**; 13.1 (C5-CH_3), 21.1 (C4-CH_3), 107.2 (C5), 113.8 (C4), 116.2 (C4'), 119.7 (C6'), 120.2 (C5'), 126.5 (C3'), 128.9 (C3), 132.0 (C2), 136.2 (C1'), 141.2 (C2'), 146.9 (C≡N). **HRMS (ESI)**; calculated for $C_{13}H_{12}N_3O_2S^+$ $[M+H]^+$ 274.0645, found 274.0643.. **HPLC**;

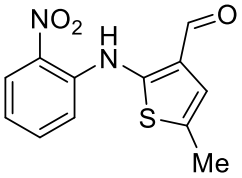
5-methyl-2-[(2-aminophenyl)amino]thiophene-3-carbonitrile (**11**) [SLM144]



Synthesised *via* method **A** 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile (4.0 mmol, 1040 mg). Compound **11** obtained as colourless plates (860 mg, 830 %).

MP; 120.2 – 121.1 °C . **^{1}H NMR (500 MHz, chloroform-*d*)**; δ 2.20 (d, J = 1.3 Hz, 3H, C5-C**H**$_3$), 3.20 – 4.00 (bs, 2H, N-**H**2), 6.19 (s, 1H, N-**H**), 6.39 (q, *J* = 1.3 Hz, C4-**H**), 6.77 (m, 2H, C5'-**H** + C6'-**H**), 7.02 (td, *J* = 1.5 + 7.7 Hz, 1H, C4'-**H**), 7.15 (dd, *J* = 1.5 + 8.0 Hz, 1H, C3'-**H**). **^{13}C NMR (125 MHz, chloroform-*d*)**; 15.1 (C5-CH_3), 87.7 (C5), 115.9 (C4), 117.1 (C5'), 119.7 (C6'), 122.3 (C4'), 124.8 (C3'), 125.3 (C3), 127.6 (C2), 127.9 (C1'), 141.0 (C2'), 160.1 (C≡N). **HRMS (ESI)**; calculated for $C_{12}H_{12}N_3S^+$ $[M+H]^+$ 230.0746, found 230.0747. **HPLC** RT = 2.33 – 2.45 mins, Area = 96.4%;

5-methyl-2-[(2-nitrophenyl)amino]thiophene-3-carbaldehyde (**12**) [SLM161]



Synthesised *via* method **B** from 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile (4.0 mmol, 1050 mg) Compound **12** obtained as a bright red needle (532 mg, 51 %).

MP 122.4 – 125.1 °C; . **^{1}H NMR (500 MHz, chloroform-*d*)**; δ 2.35 (d, *J* = 1.3 Hz, 3H, C5-C**H**$_3$), 6.77 (q, *J* = 1.3 Hz, 1H, C4-**H**), 7.02 (ddd, *J* = 1.3 + 7.1 + 8.5 Hz, 1H, C4'-**H**), 7.55 (ddd, J = 1.6 + 7.1 + 8.5 Hz, 1H, C5'-**H**), 7.79 (dd, *J* = 1.3 + 8.5 Hz, 1H, C6'-**H**), 8.20 (dd, *J* = 1.6 + 8.5 Hz, 1H, C3'-**H**), 9.73 (s, 1H, N-**H**), 12.29 (s, 1H, **H**C=O). **^{13}C NMR (125 MHz, chloroform-*d*)**; 14.9 (C5-CH_3), 117.5 (C5), 121.3 (C4), 123.6 (C4'), 123.7 (C5'), 126.2 (C6'), 126.9 (C3'), 135.5 (C3), 136.2 (C2), 137.5 (C1'), 152.2 (C2'), 184.8 (C=O). **HRMS (ESI)**; calculated for $C_{12}H_{11}N_3O_2S^+$ $[M+H]^+$ 274.0645, found 263.0483. **HPLC** RT = 3.30 – 3.40 mins, Area = 98.2%;

5.3 Single Crystal XRD

For a summary of CSD deposition numbers and space group information refer to Table 7 on page 128.

A suitable single crystal was selected and immersed in inert oil. The crystal was then mounted to a goniometer head on an Agilent SuperNova X-ray

diffractometer fitted with an Atlas area detector and a kappa-geometry 4- circle goniometer, using mirror monochromated Cu-Kα (λ = 1.54184 Å) radiation. The crystal was cooled to 100 K by an Oxford cryostream low temperature device. The full data set was recorded and the images processed using CrysAlis Pro[103]. Structure solution by intrinsic phasing was achieved through the use of SHELXT[104] and the structural model refined by full matrix least squares on F^2 using SHELXL[105], interfaced through Olex 2[106]. Molecular graphics were plotted, editing of CIFs and construction of tables of bond lengths and angles were achieved using Olex2. Unless otherwise stated, hydrogen atoms were placed using idealised geometric positions (with free rotation for methyl groups), allowed to move in a “riding model” along with the atoms to which they were attached, and refined isotropically. All structural data were collected by the researcher under the supervision of Christopher Pask. Structural solutions were performed by the researcher and Christopher Pask.

5.4 Polymorph Screen

For a summary of experimental parameters for the polymorph screen please refer to Figure 139 on page 187. For details of the results of the polymorph screen, please refer to Table 13 and Table 14 on pages 190 and 192

Chapter 4 described a polymorph screen designed to explore the polymorphic landscape of the 12 ROY analogues synthesised in Chapter 3. The focus of this screen was to identify any easily accessible polymorphic forms of the 12 ROY analogues, in an attempt to verify the status of the ROY scaffold as a polymorphophore. Each ROY analogue synthesised in Chapter 3 was subject to an individual polymorph screen using the setup described below.

100-well StarStore polypropylene storage boxes were used to contain all screening experiments for a single compound, with Agilent 2 mL glass screw-top vials used to contain the individual crystallisation experiments. Vials and polypropylene storage boxes were purchased from Science Warehouse. 12 unique crystallisation conditions were used in duplicate for each ROY analogue, resulting in 24 crystallisation experiments per analogue.

The conditions of the twelve crystallisation experiment were as follows: acetone, acetonitrile, chloroform, ethyl acetate, acetone (reduced evaporation rate), acetonitrile (reduced evaporation rate), chloroform (reduced evaporation rate), ethyl acetate (reduced evaporation rate), 10:1 acetone/water, 10:1 acetonitrile/methanol, 10:1 chloroform/ethanol and 10:1 ethyl acetate/hexane. Reduced evaporation rate was achieved by covering the vial with aluminium foil, whilst all remaining experiments proceeded uncovered.

Crystallisation experiments were set up as follows: 10 mg of ROY analogue was weighed into a 2 mL glass screw-top vial. 2mL of corresponding solvent was added to the vial using an accu-pipette, before the lid was screwed on and the contents shaken vigorously to mix. After mixing, the cap was removed and the vial placed in the corresponding slot in the polypropylene storage box. Aluminium foil was used to cover the top of reduced evaporation rate experiments. Once setup, the polypropylene storage box was stored in a fumehood and left for ~ a week to allow crystallisations to occur. All screening experiments were performed at room temperature.

5.4.1 Polymorph Screen Detection

Once all crystallisation experiments for a given compound were complete the vials were inspected for signs of new polymorphic forms. As discussed in Chapter 4, the colour polymorphism displayed by ROY and some of its analogues was used as the initial test for identifying new polymorphic forms. The crystal observed within the 2 mL glass vial was compared to a sample of crystal of the ROY analogue characterised in Chapter 3, and any changes in colour or habit were identified as potential new polymorphic forms.

Vials containing potential new polymorphic forms were crushed into a fine powder using mortar and pestle, and this powder was analysed using (machine make and model) PXRD using a Cu-β radiation source at 1.58 Å wavelength. Each powdered sample was analysed for 40 minutes in order to produce an polymorph powder pattern.

The CSD powder pattern comparison tool/script (accessible here) was used to compare the polymorph powder pattern to the simulated powder pattern of the crystal of the corresponding ROY analogue characterised in Chapter 3 in order to determine the nature of the crystals obtained via the polymorph screen. Typically a threshold of < 95% powder pattern match is indicative of a new polymorphic form. However, using the CSD script to compare simulated PXRD patterns of 11 CB (characterised in Chapter 3) with experimentally obtained PXRD patterns of 11 CB yielded a match of only 93%. As such, a threshold of 92% match was used as a means to detect new polymorphic forms.

6.1 NMR Spectra applying to Chapter 3

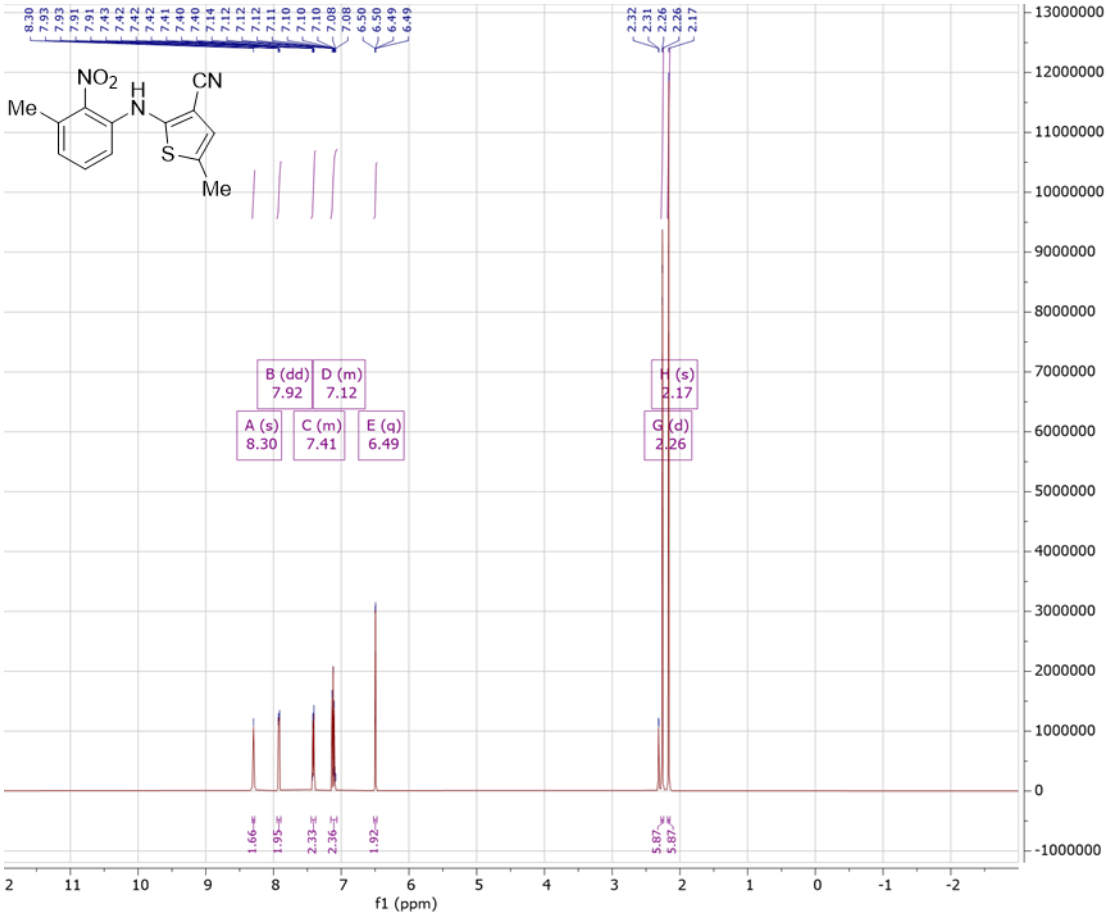


Figure 142; 1H NMR spectrum of SLM143/Derivative 1

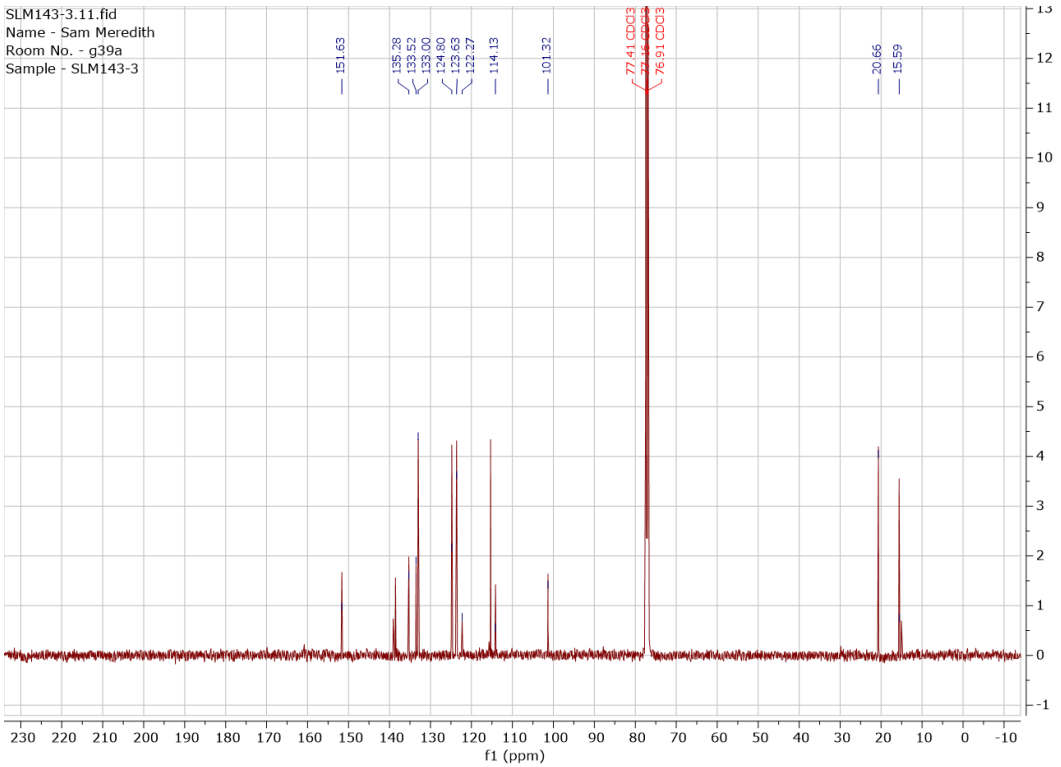


Figure 143; 13C NMR spectrum of SLM143/Derivative 1

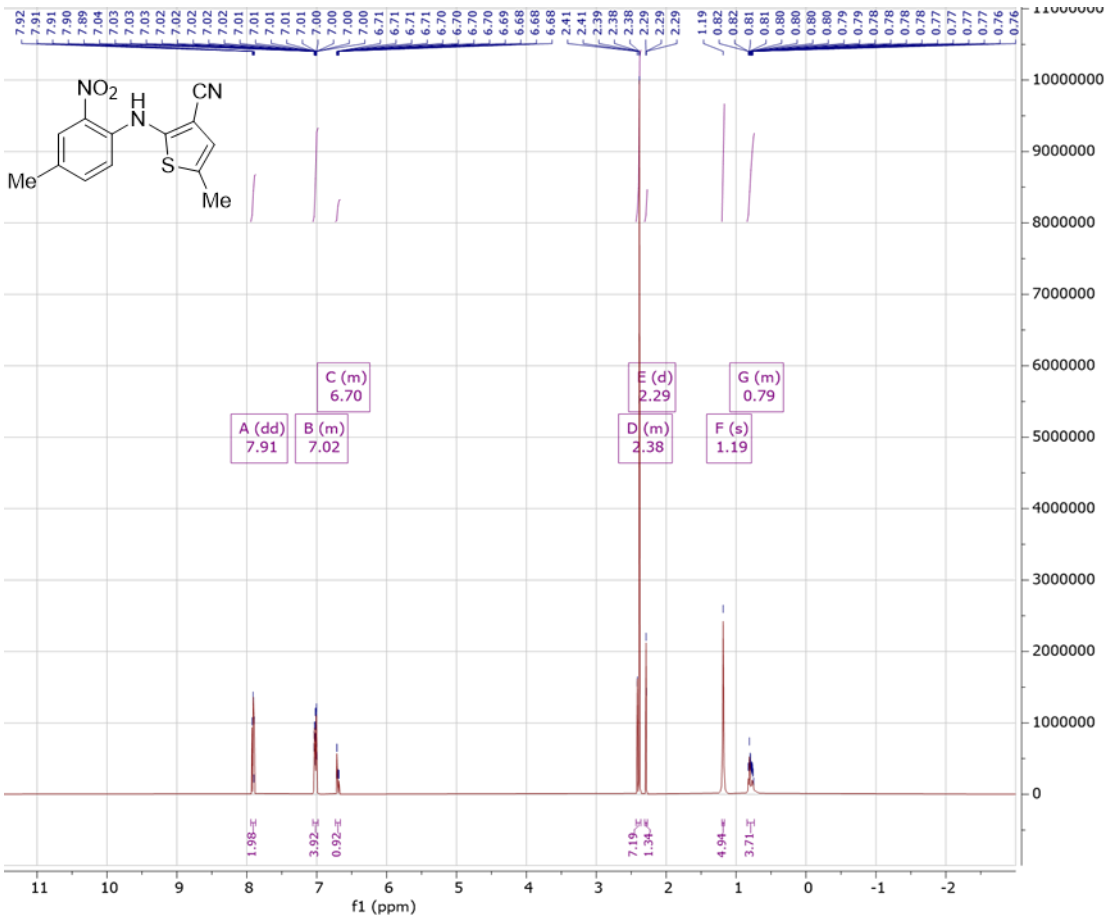


Figure 144; 1H NMR spectrum of SLM166/Derivative 2

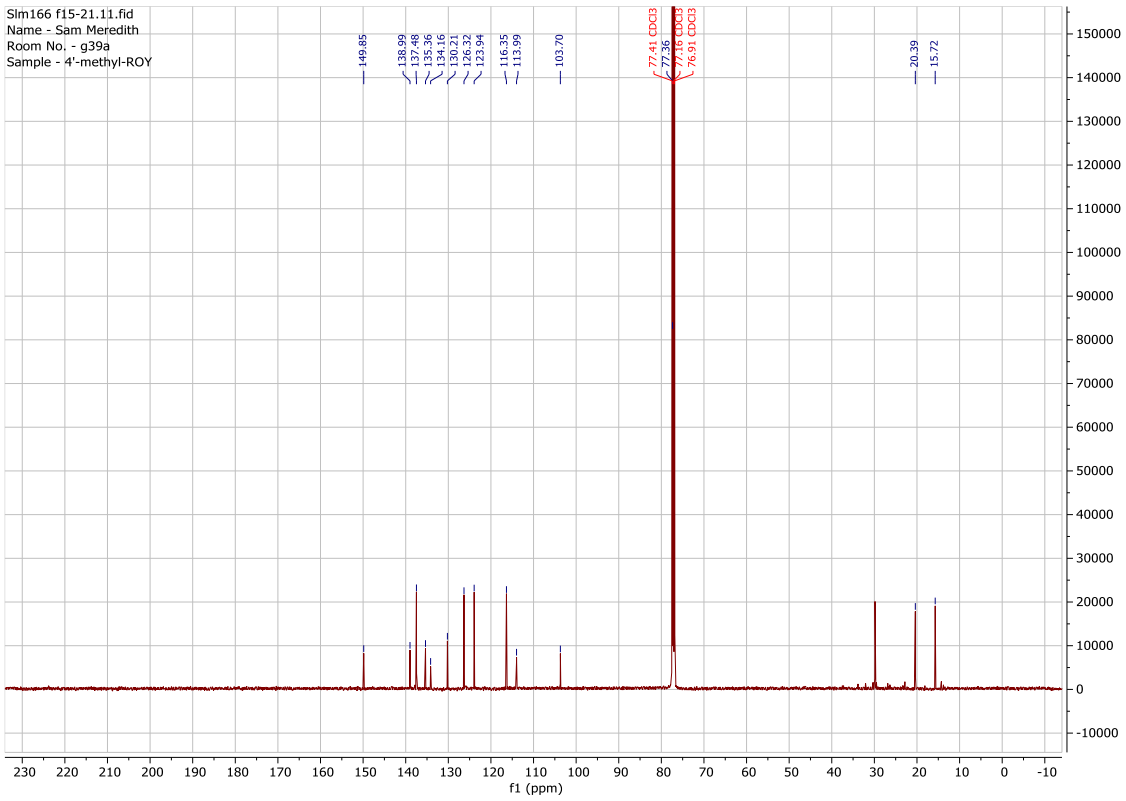


Figure 145; 13C NMR spectrum of SLM166/Derivative 2

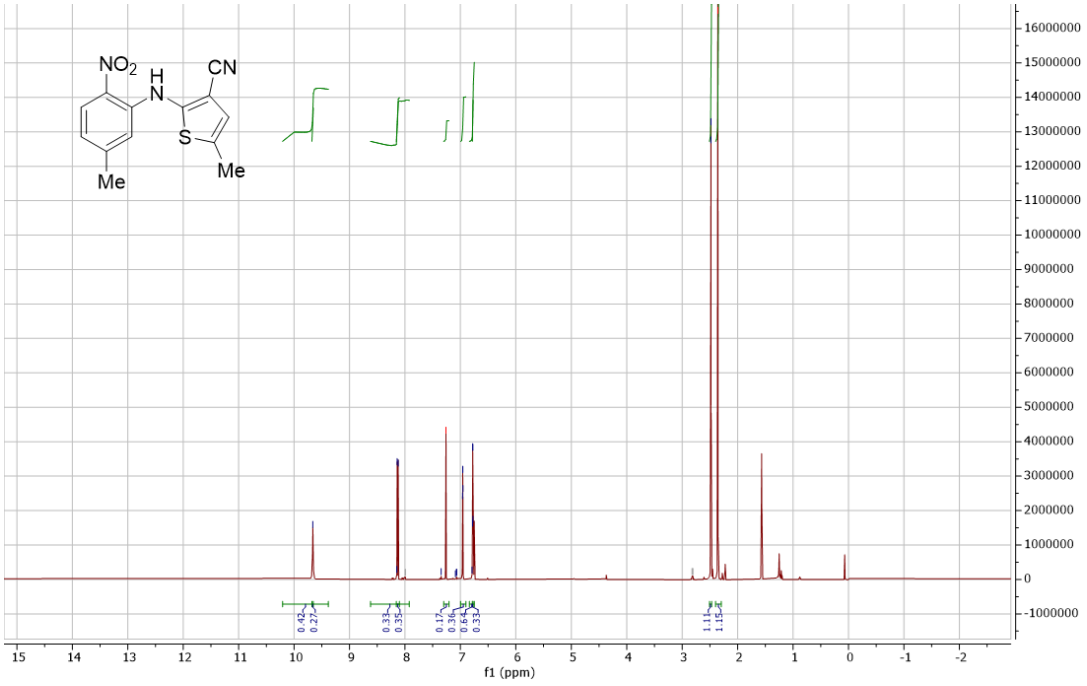


Figure 146; 1H NMR spectrum of SLM168/Derivative 3

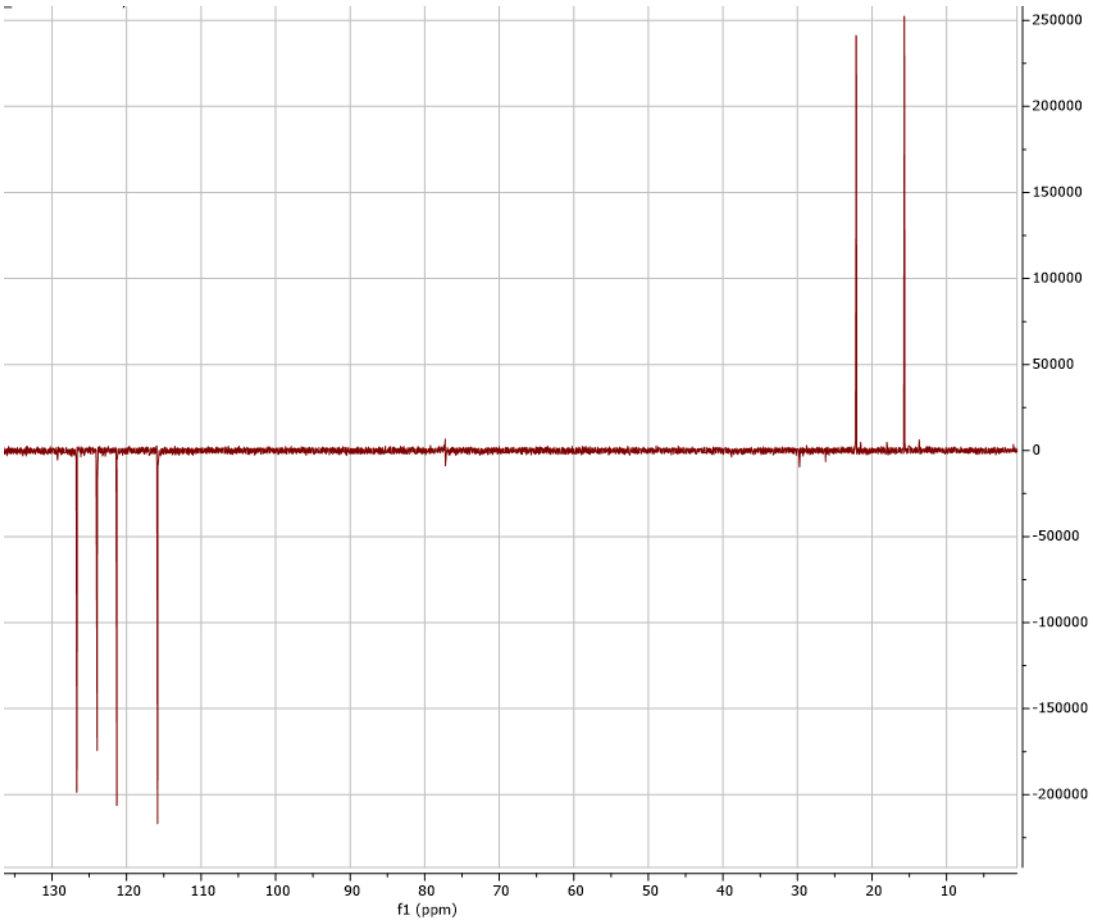


Figure 147; DEPT NMR spectrum of SLM168/Derivative 3

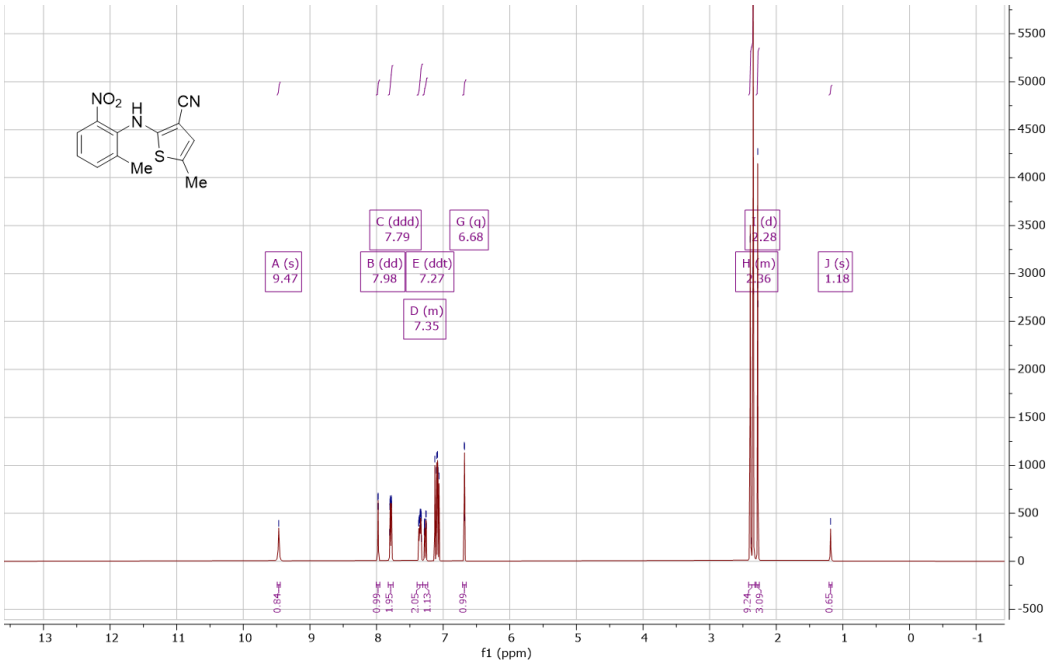


Figure 148; 1H NMR Spectrum of SLM142/Derivative 4

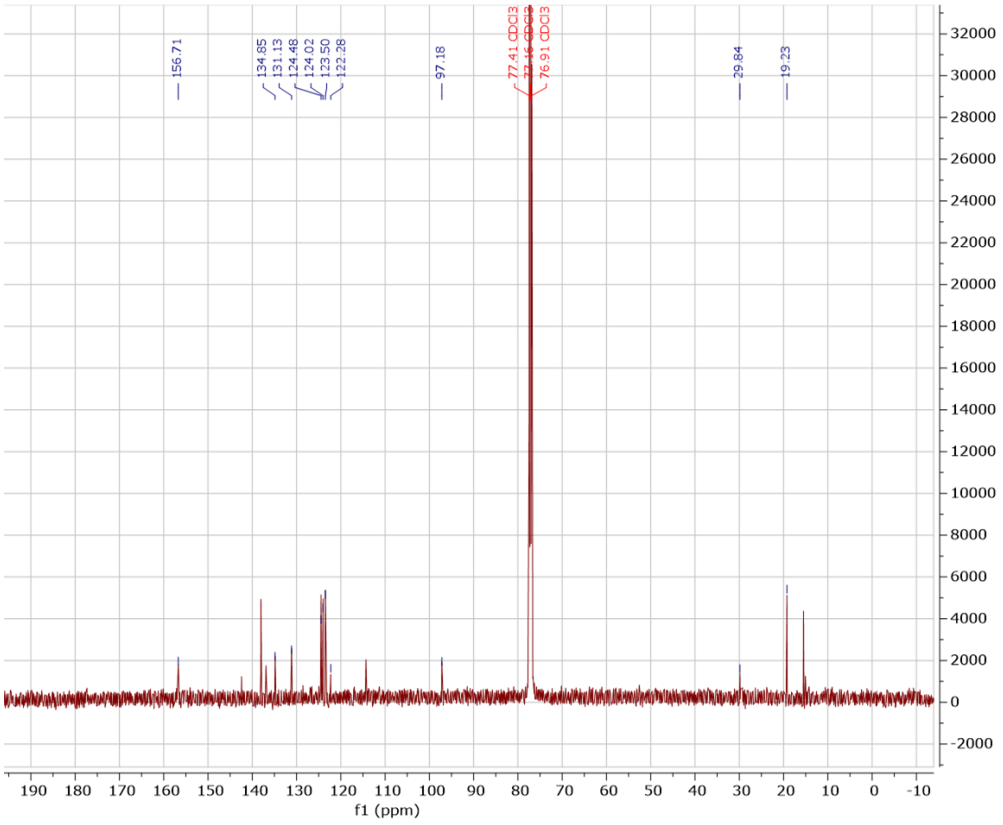


Figure 149; 13C NMR spectrum of SLM142/Derivative 4

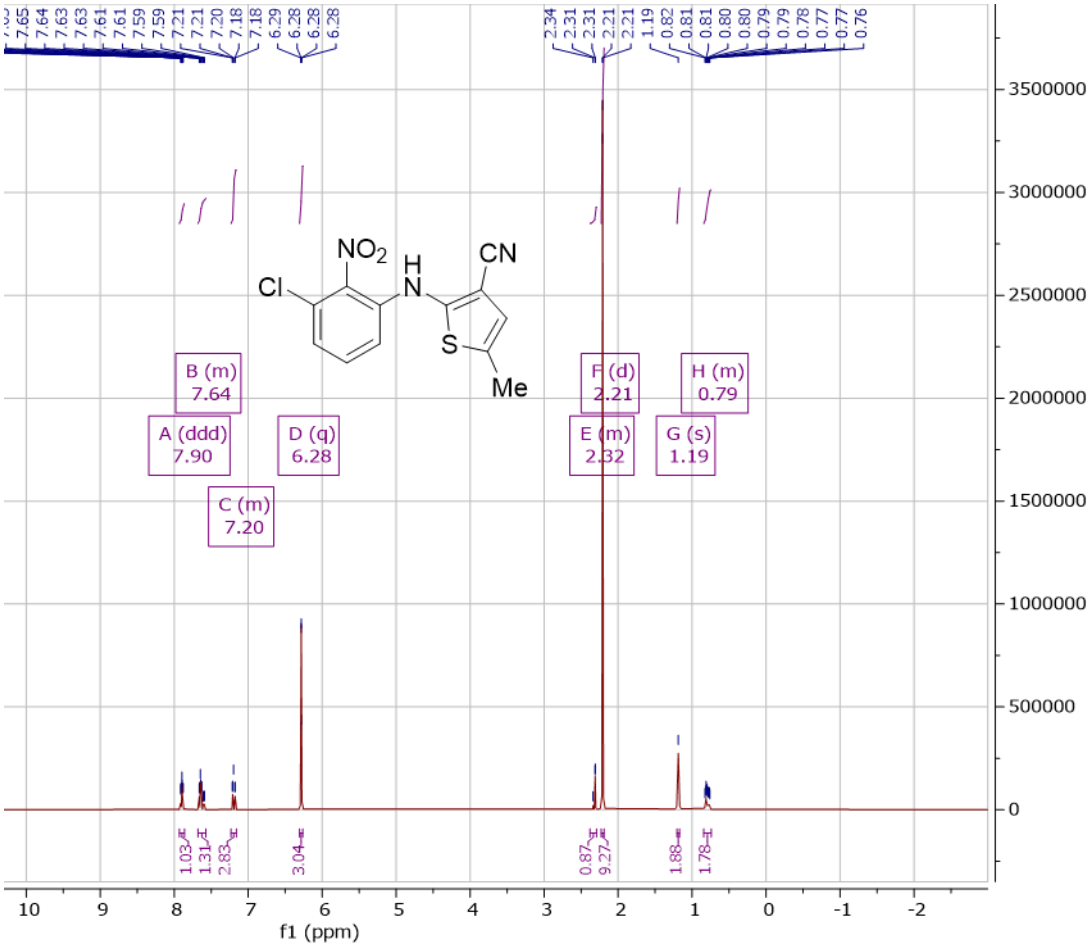


Figure 150; 1H NMR spectrum of SLM96/Derivative5

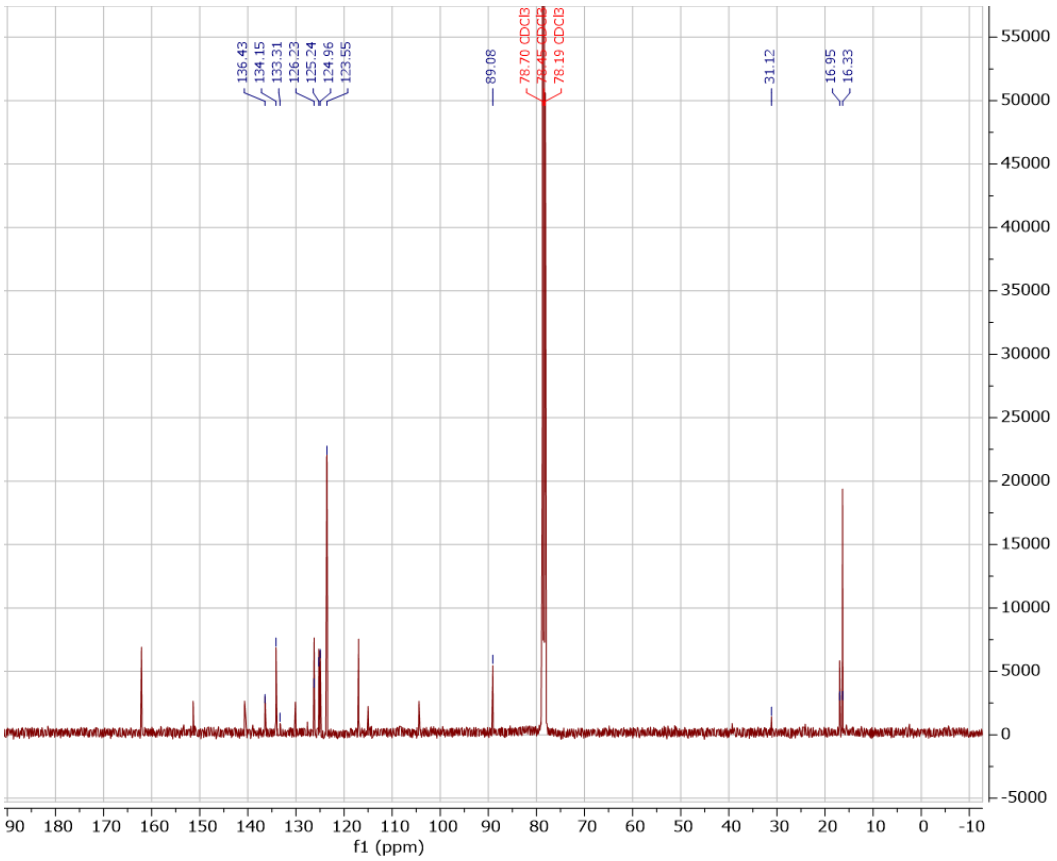


Figure 151; 13C NMR spectrum of SLM96/Derivative 5

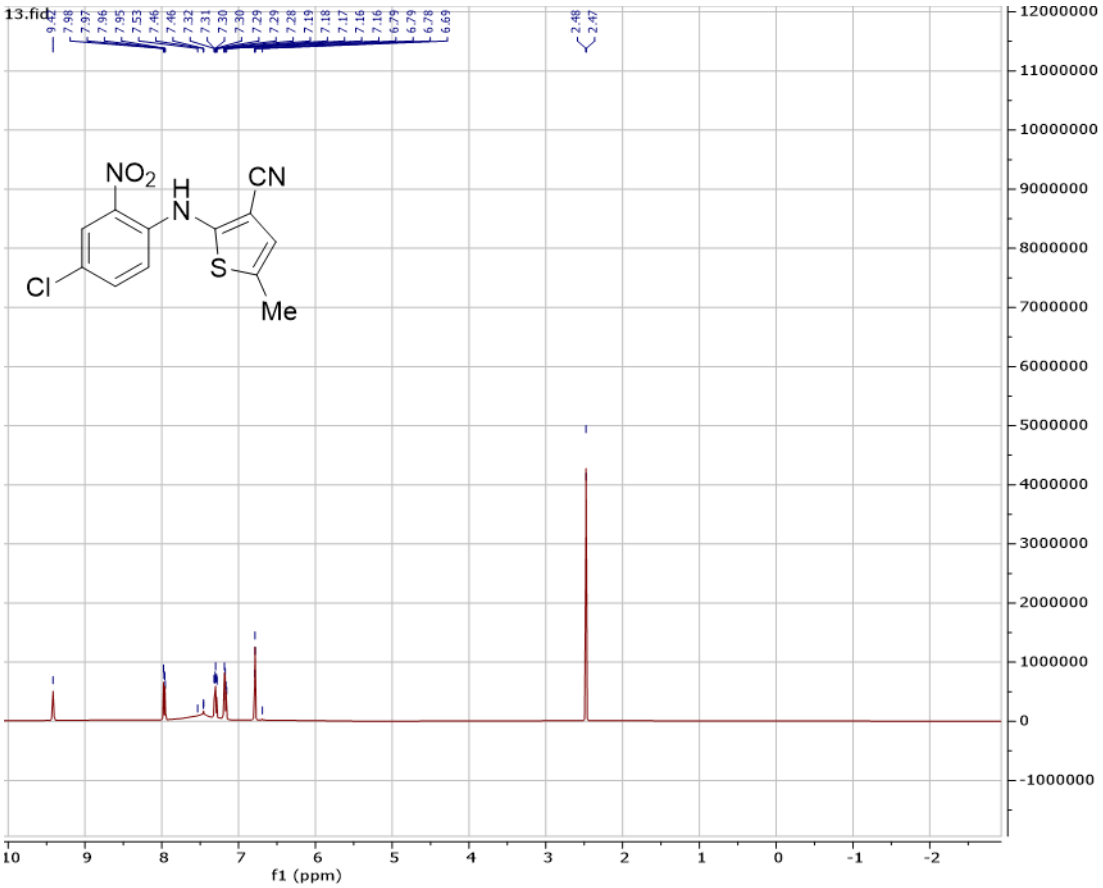


Figure 152; 1H NMR spectrum of SLM169/Derivative 6

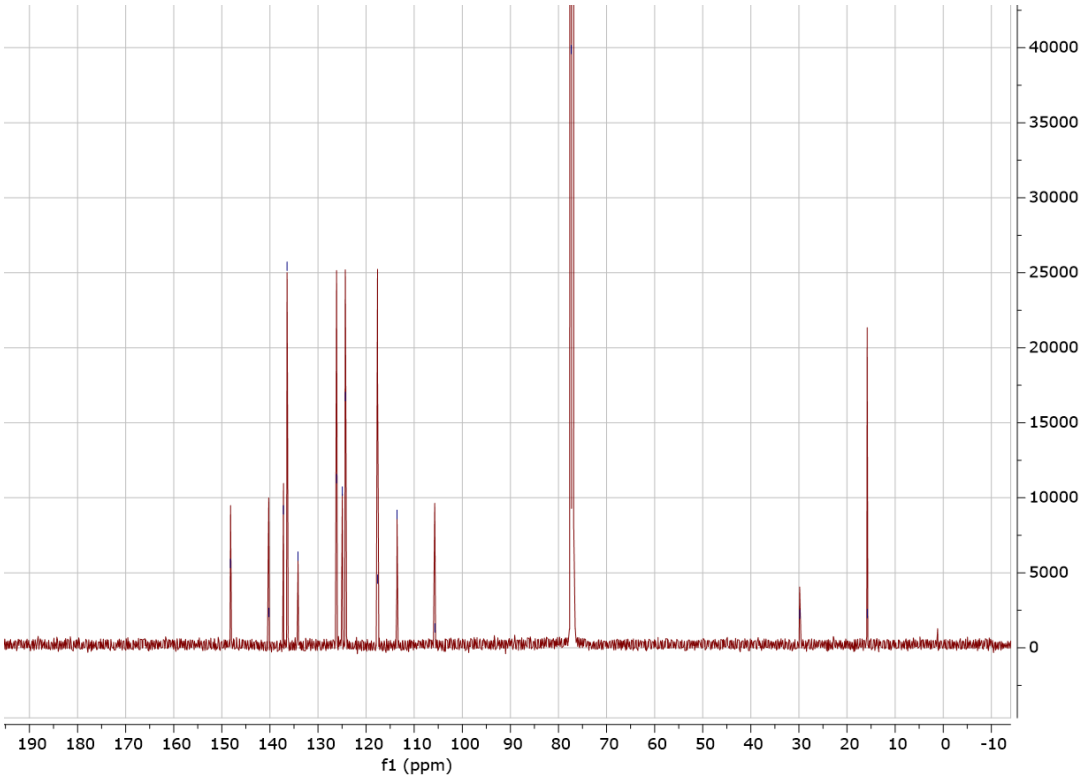


Figure 153; 13C NMR spectrum of SLM169/Derivative 6

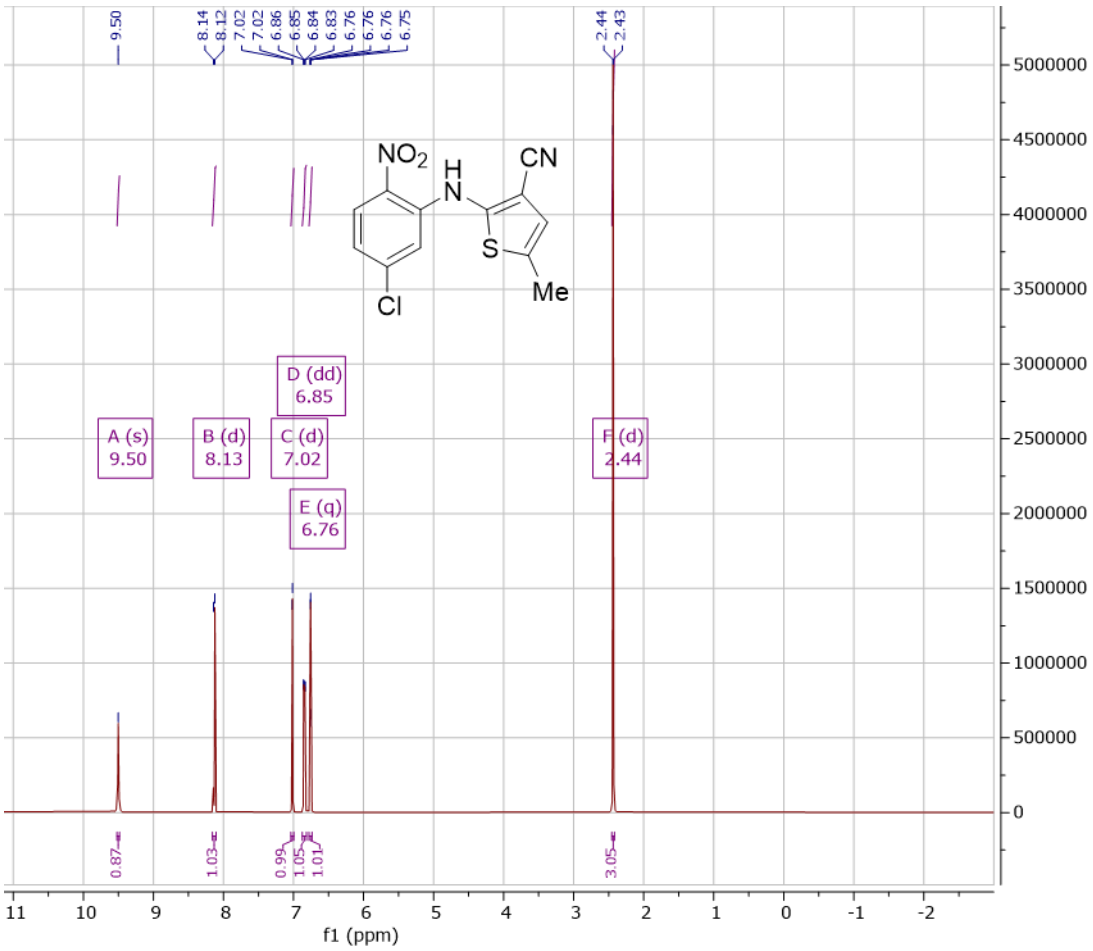


Figure 154; 1H NMR spectrum of SLM94/Derivative 7

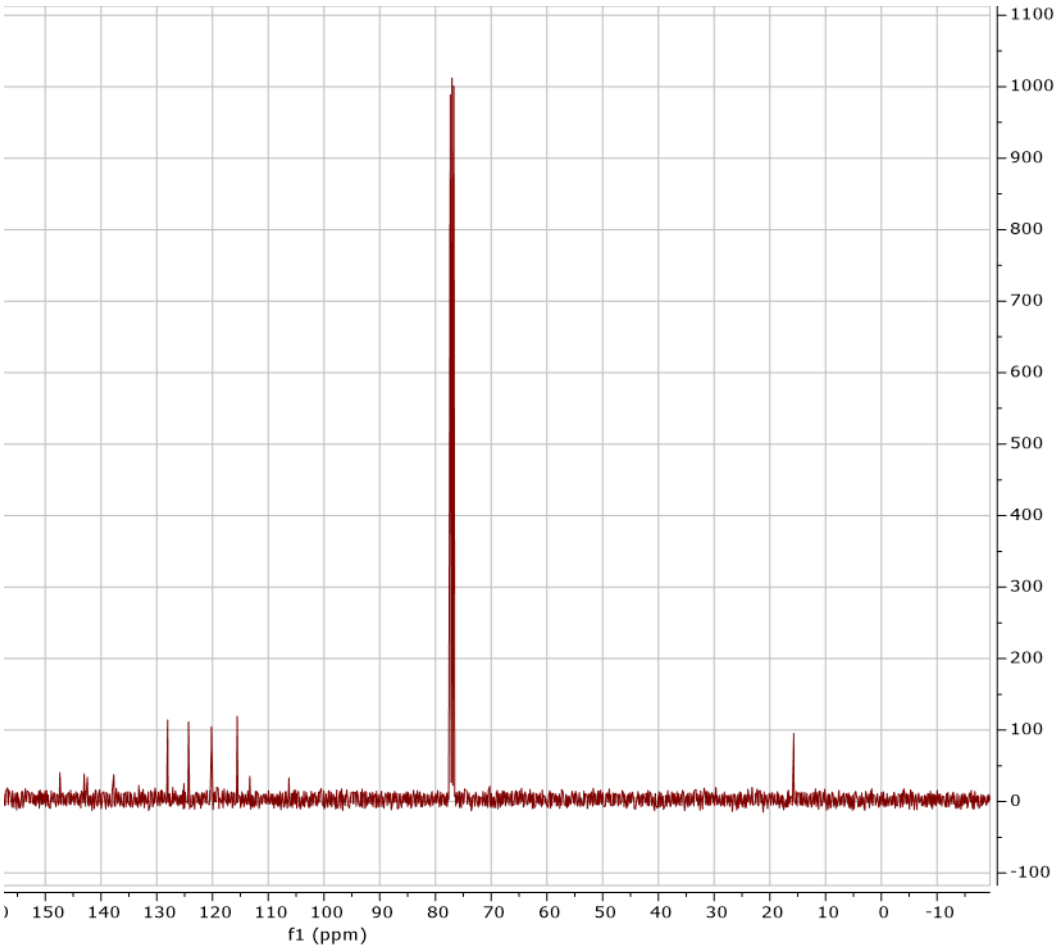


Figure 155; 13C NMR spectrum of SLM94/Derivative 7

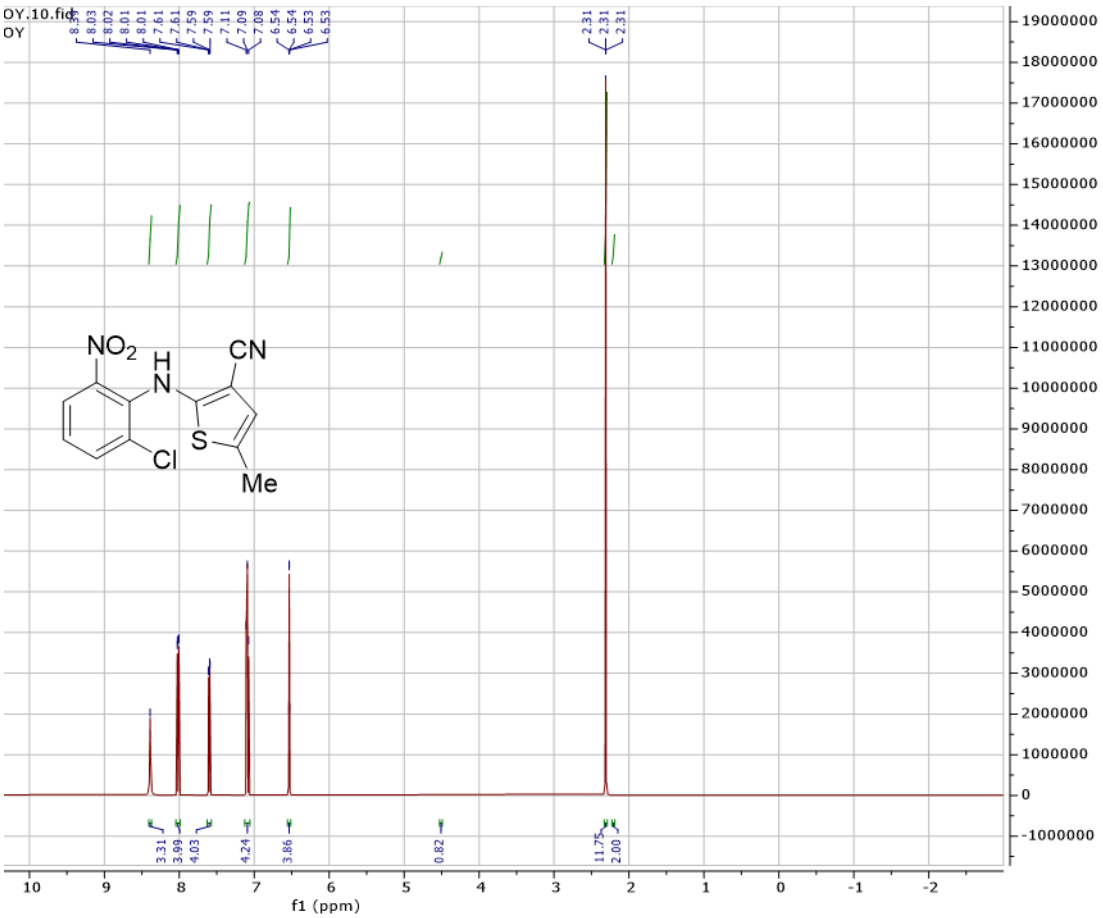


Figure 156; 1H NMR spectrum of SLM165/Derivative 8

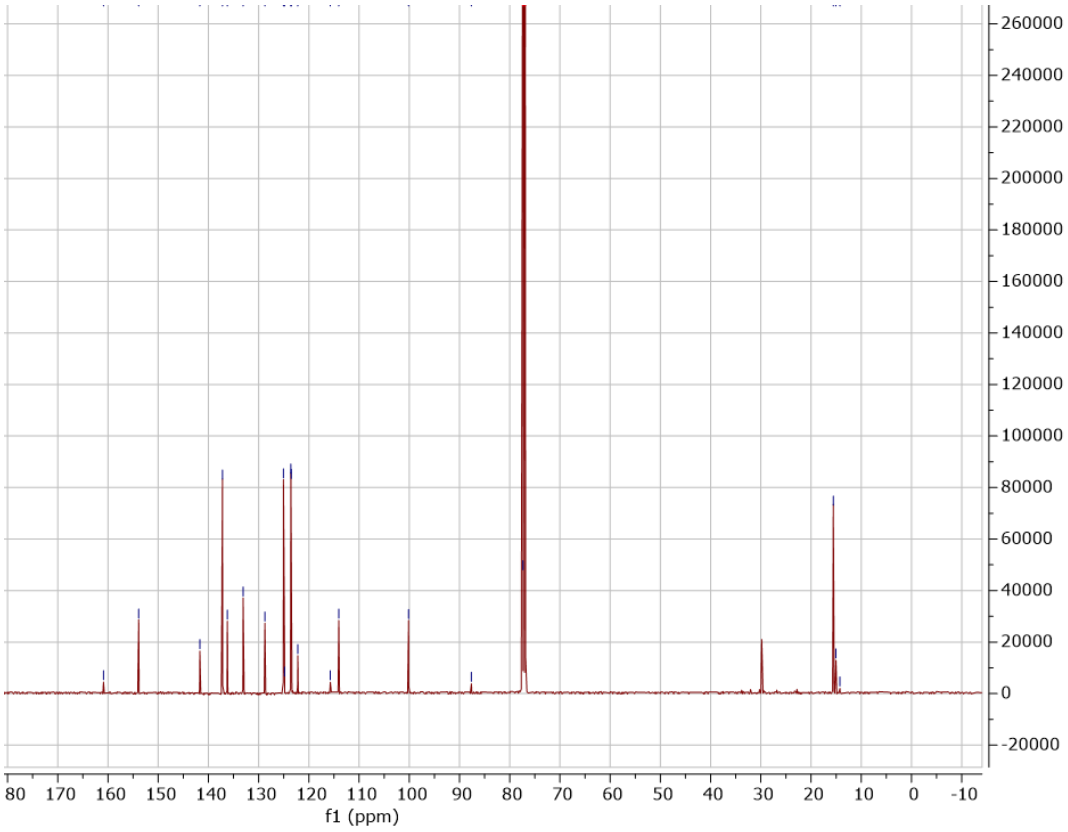


Figure 157; 13C NMR spectrum of SLM165/Derivative 8

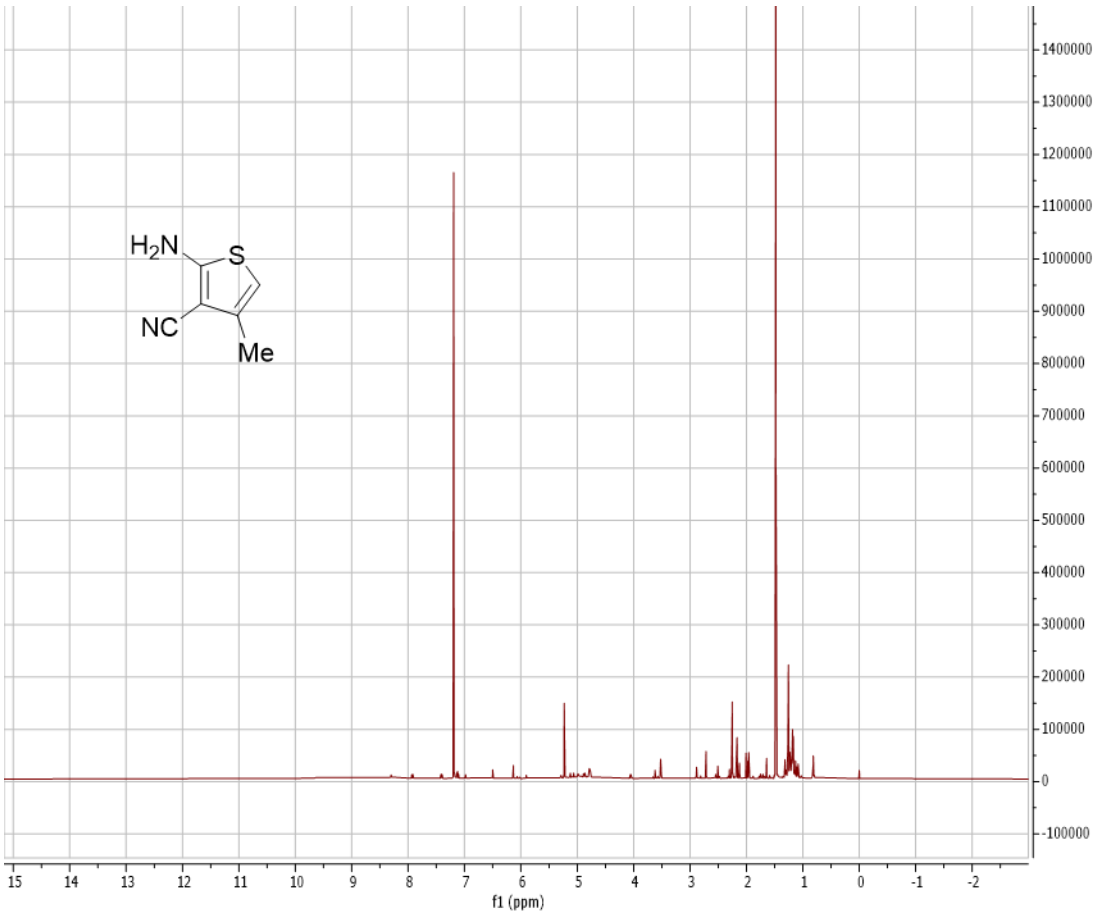


Figure 158; 1H NMR spectrum of SLM113/Derivative 9a

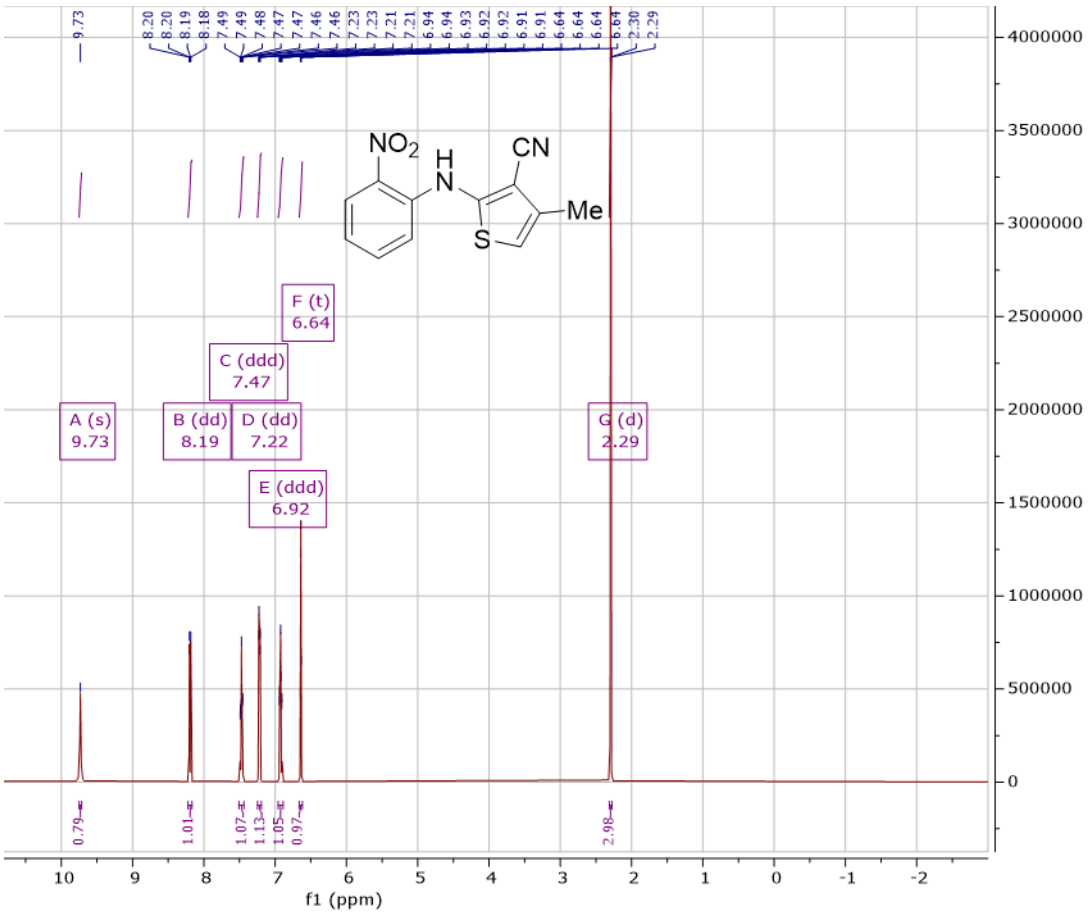


Figure 159; 1H NMR spectrum of SLM141/Derivative 9

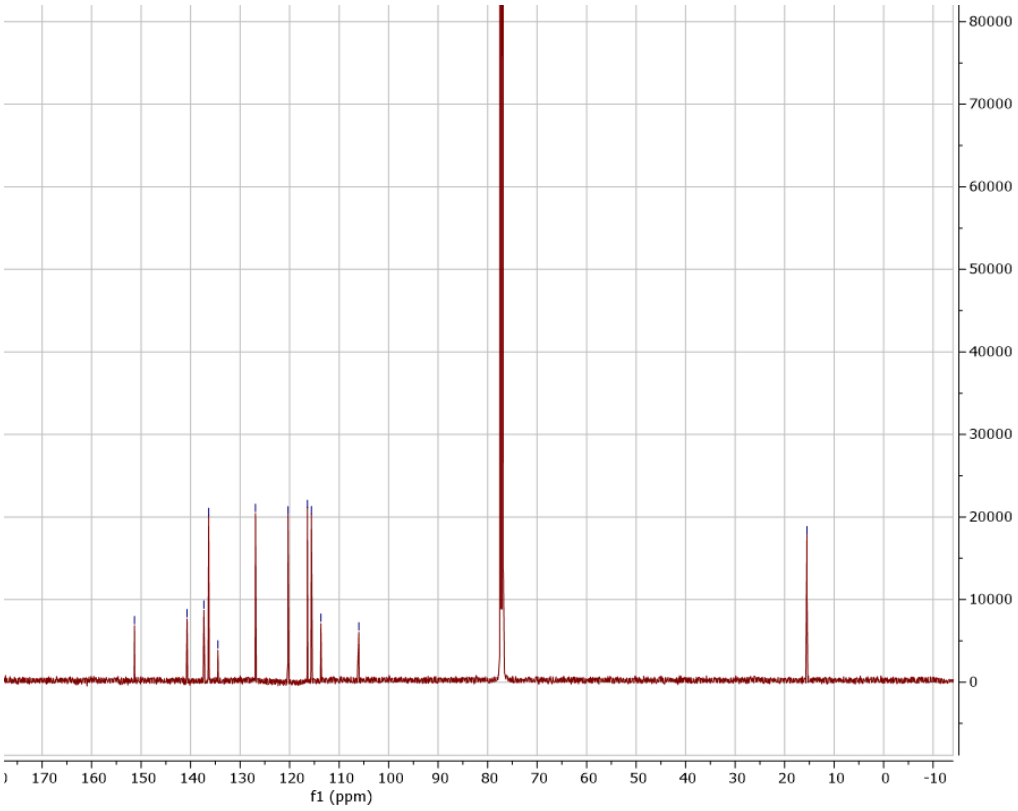


Figure 160; 13C NMR Spectrum for Derivative 9/SLM141

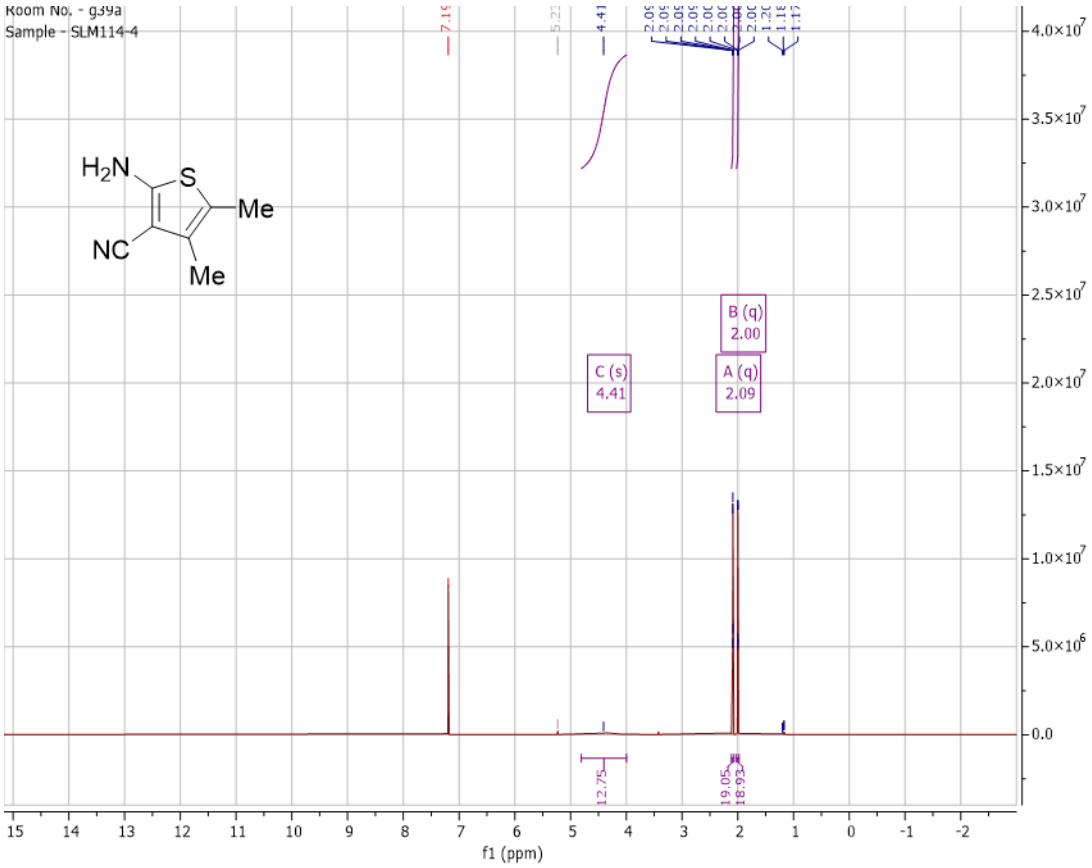


Figure 161; 1H NMR spectrum of SLM114/Derivative 10a

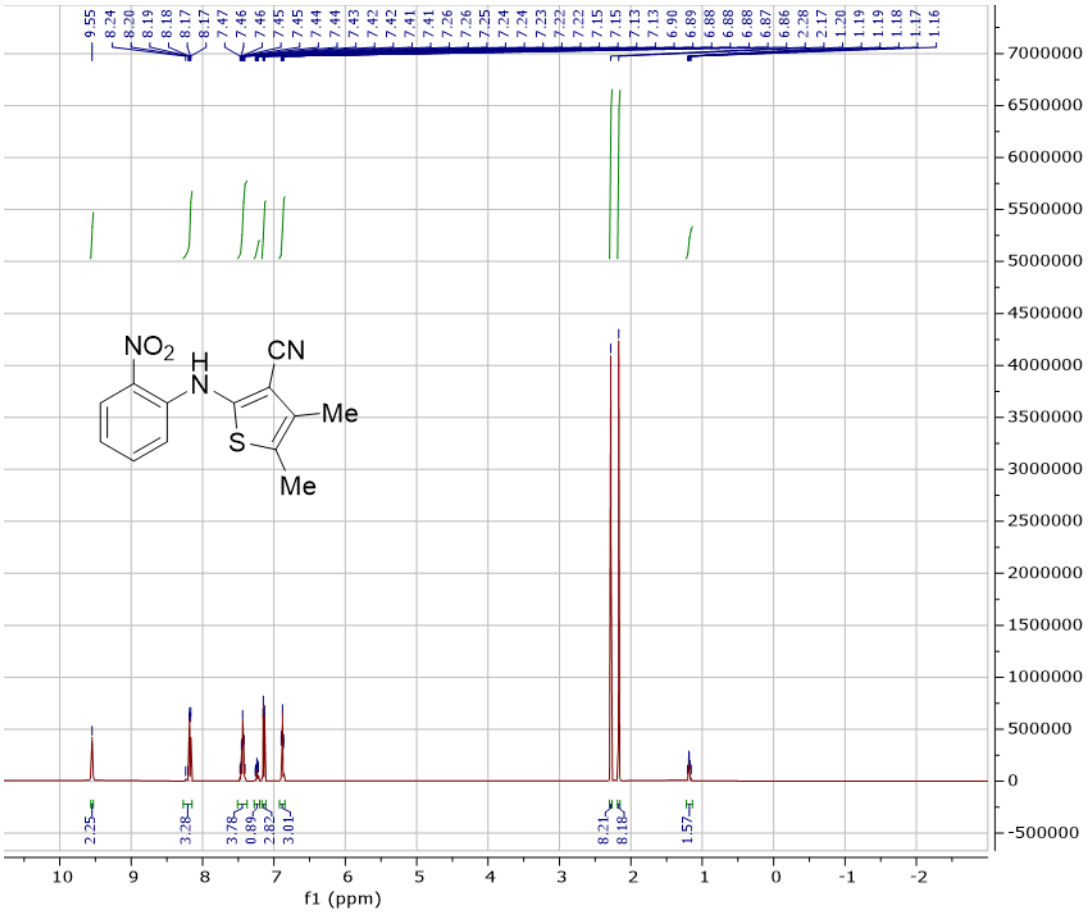


Figure 162; 1H NMR spectrum of SLM140/Derivative 10

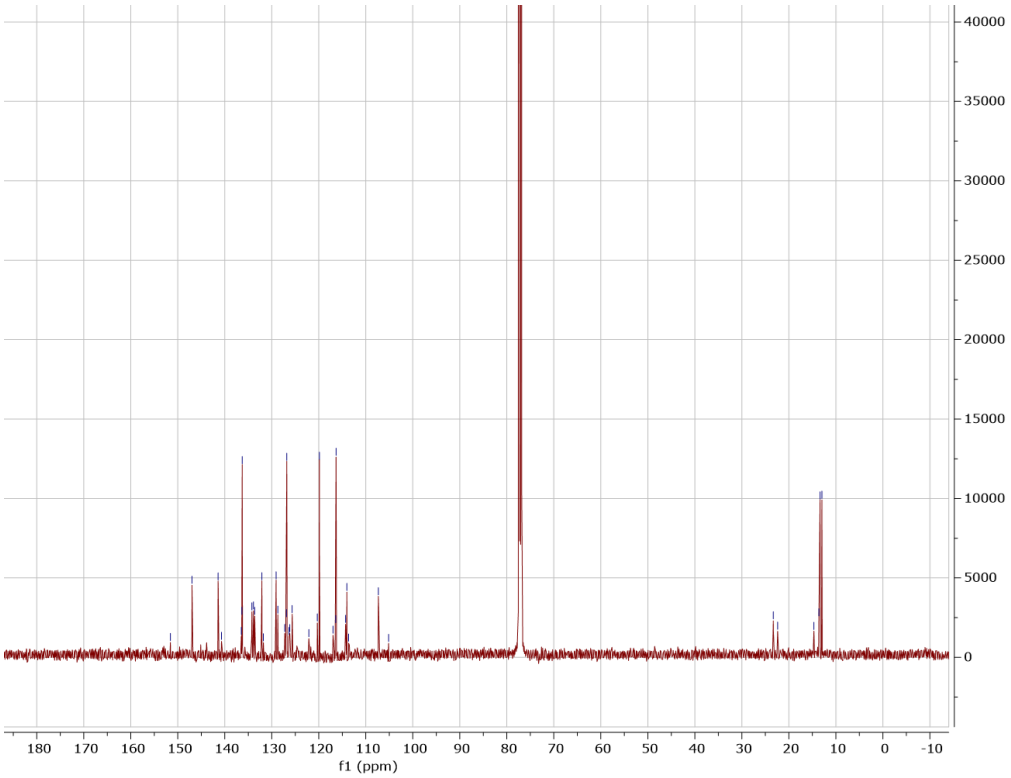


Figure 163; 13C NMR Spectrum of SLM 140/Derivative 10

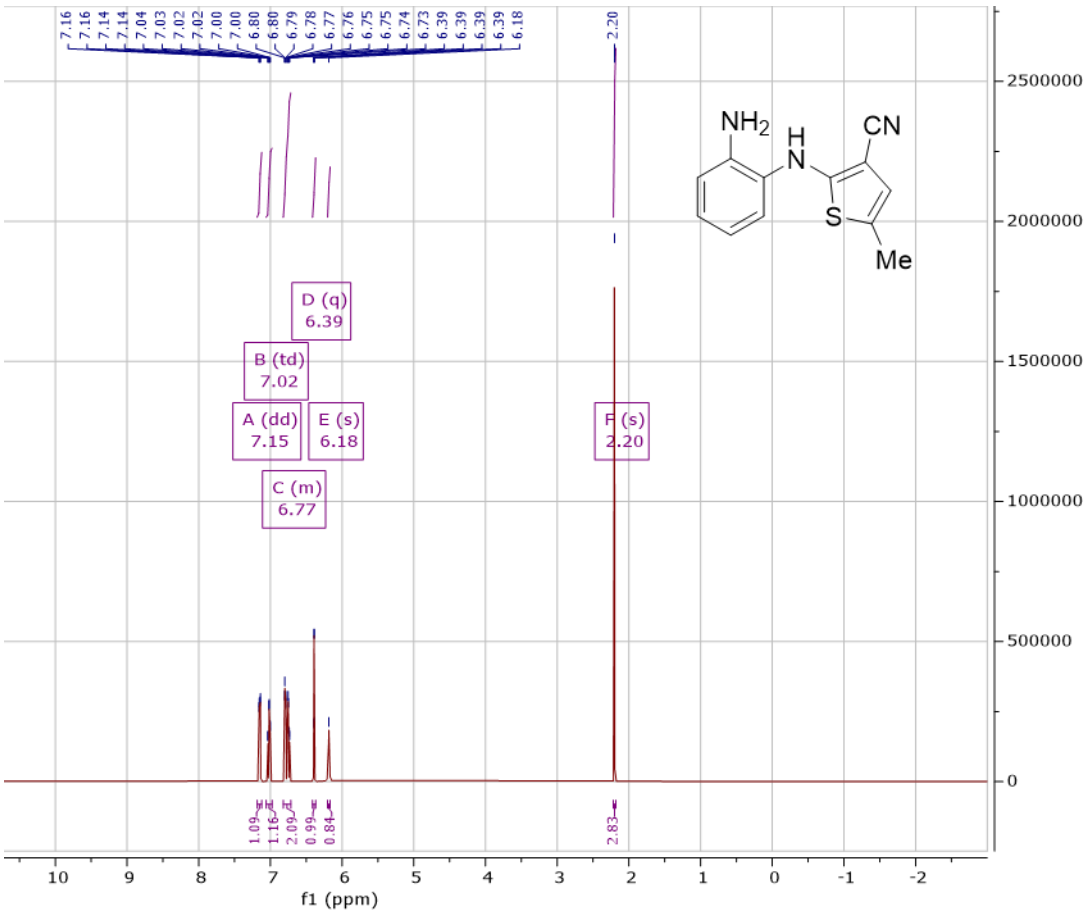


Figure 164; 1H NMR spectrum of SLM106/Derivative 11

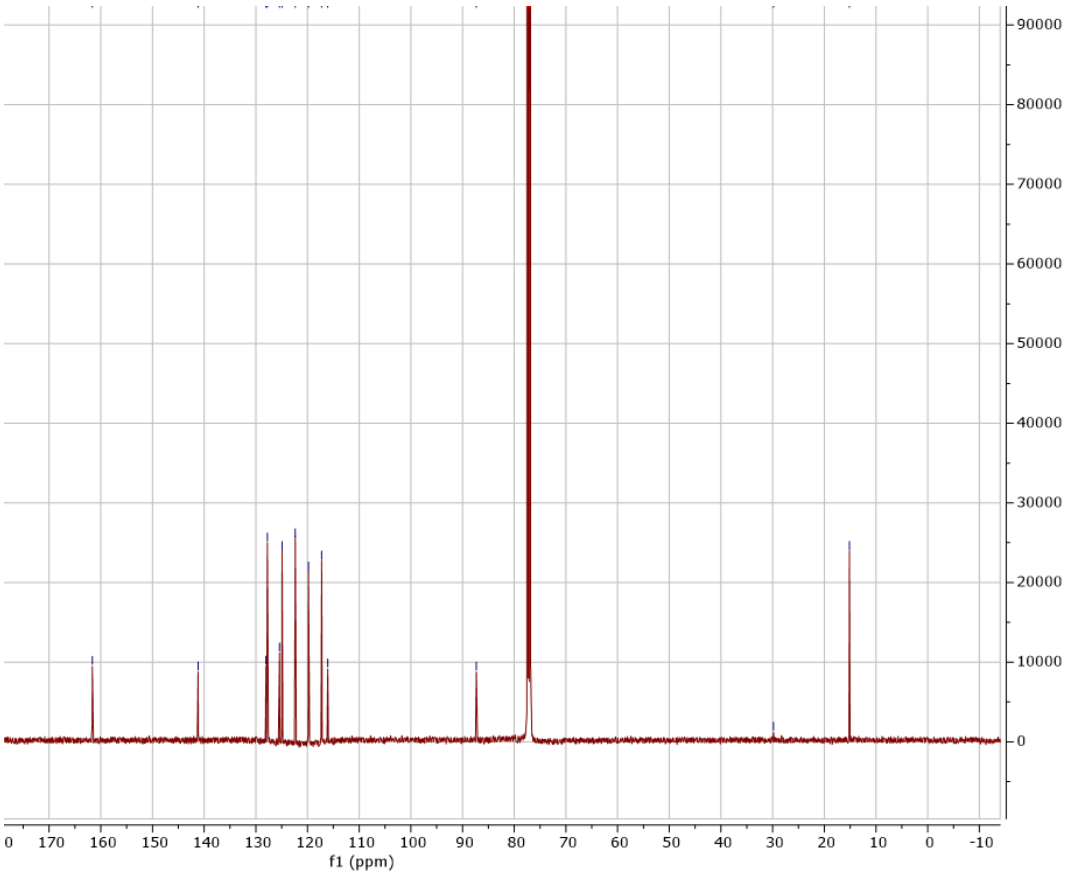


Figure 165; 13C NMR spectrum of SLM106/Derivative 11

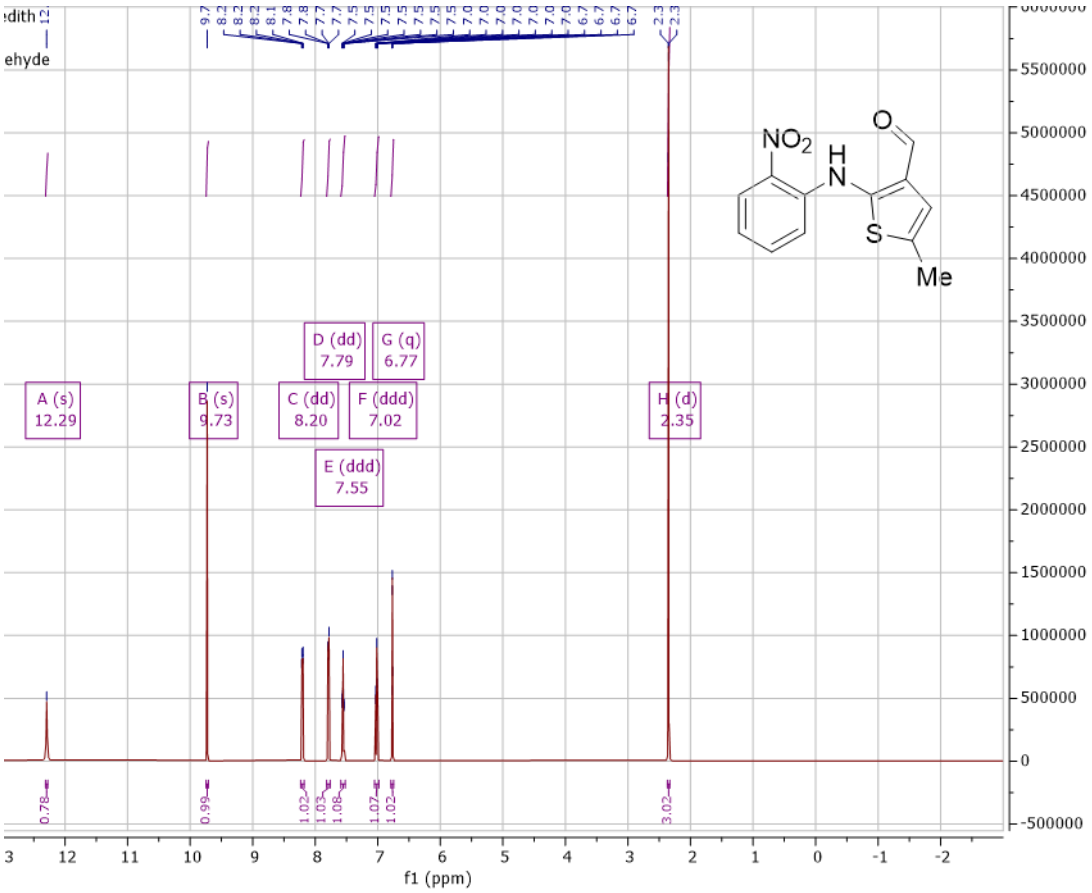


Figure 166; 1H NMR spectrum of SLM161/Derivative 12

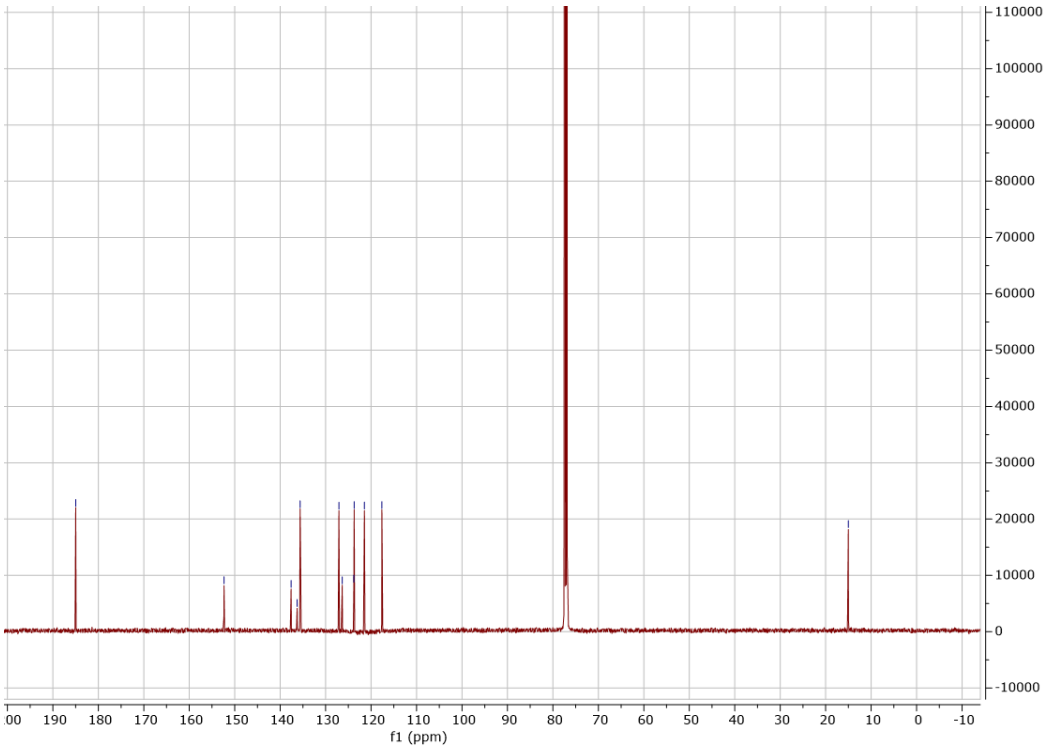


Figure 167; 13C NMR Spectrum for SLM161/Derivative 12

6.2 HPLC chromatograms relating to Chapter 3

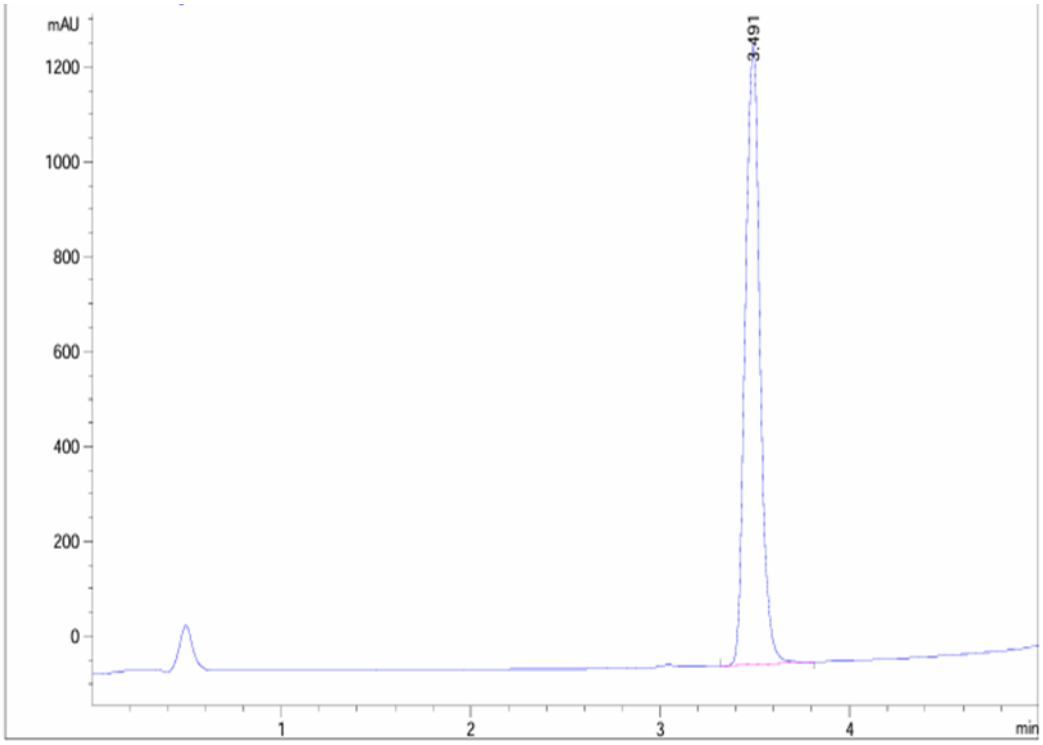


Figure 168; HPLC trace for SLM143/Derivative 1

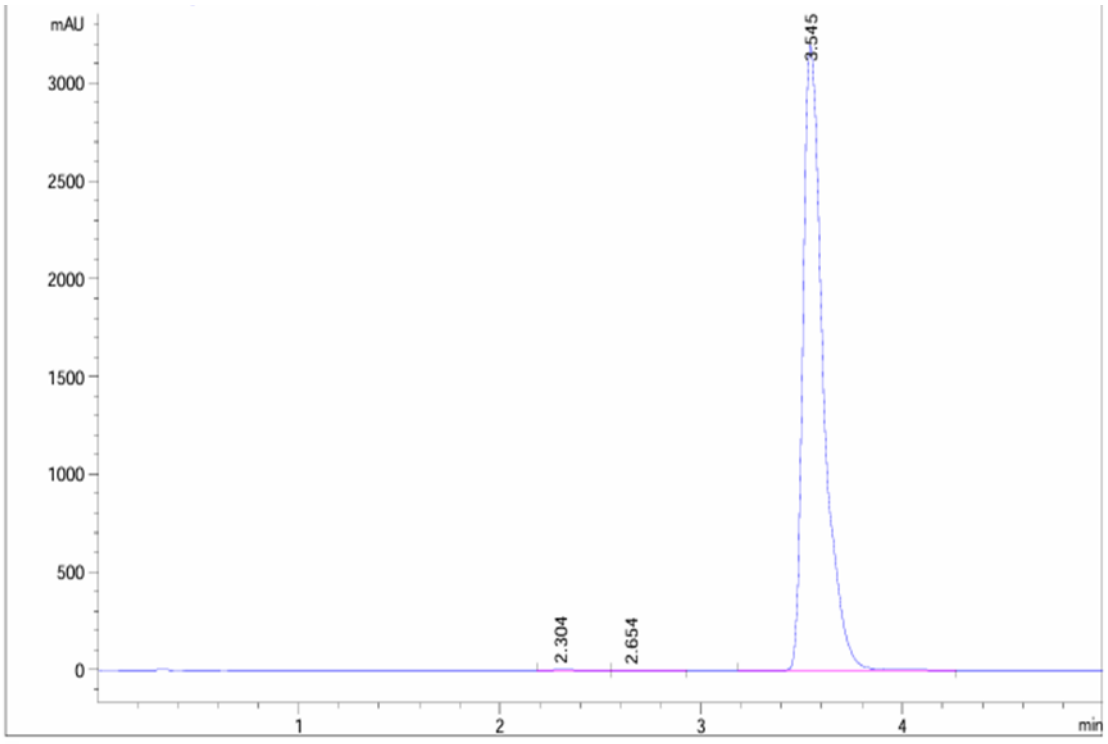


Figure 169; HPLC trace for SLM166/Derivative 2

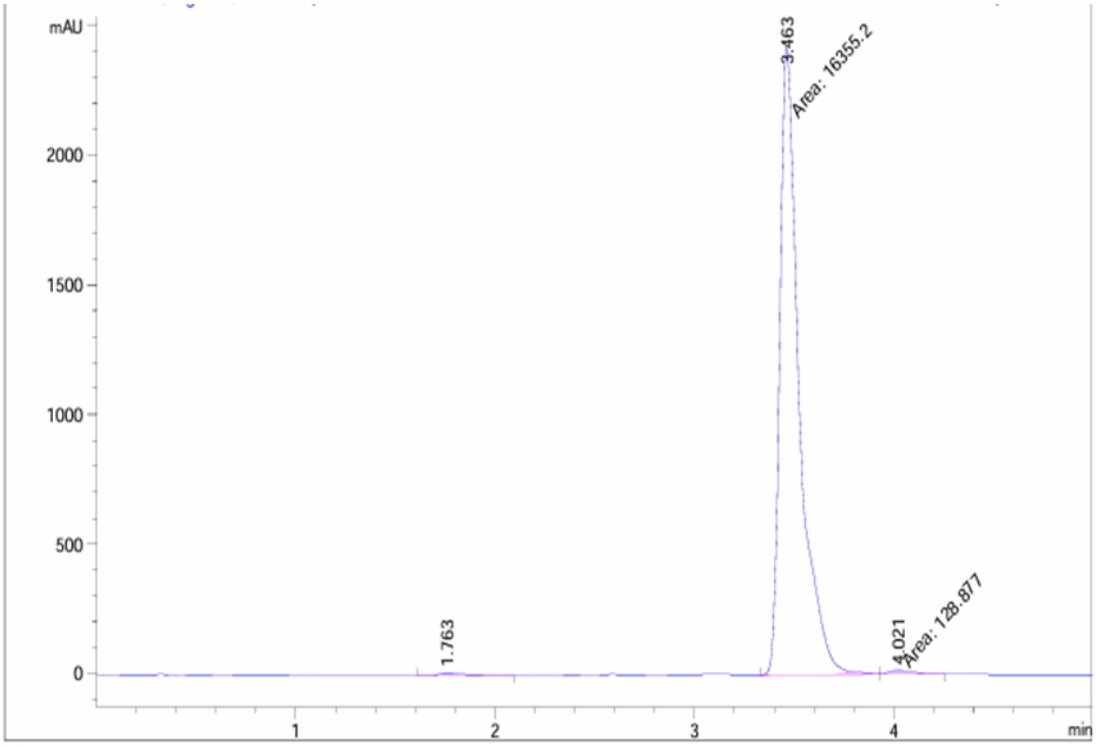


Figure 170; HPLC Trace for SLM 168/Derivative 3

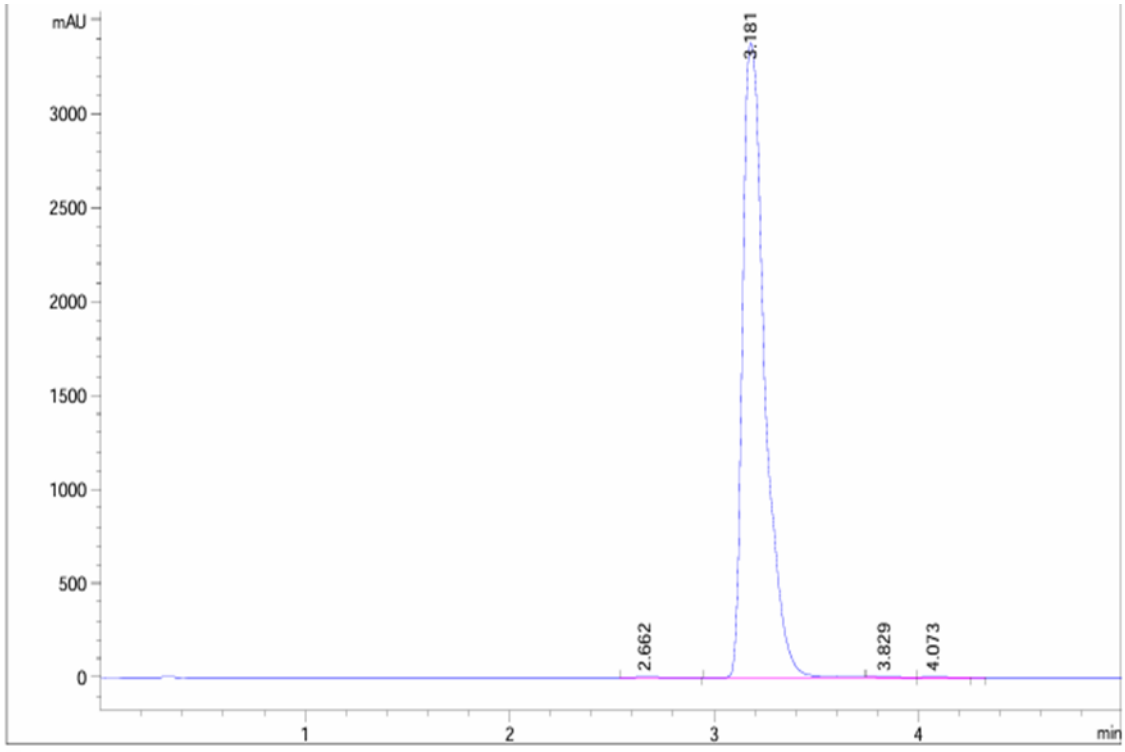


Figure 171; HPLC Trace for SLM142/Derivative 4

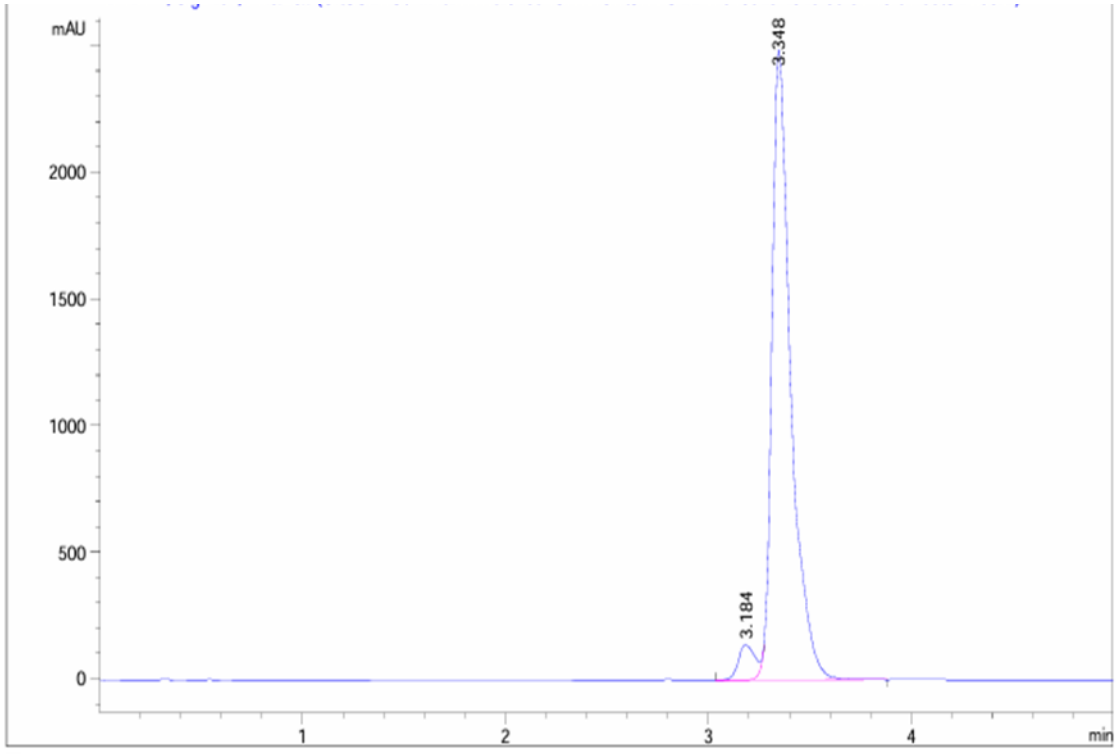


Figure 172; HPLC trace for SLM96/Derivative 5

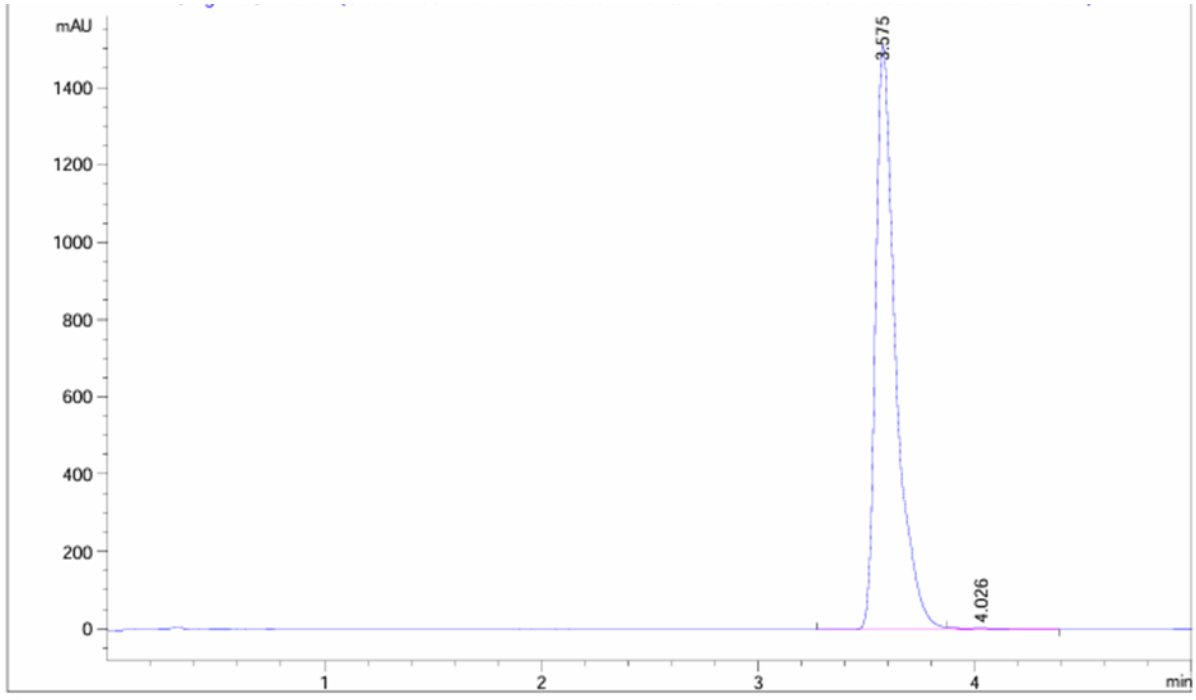


Figure 173; HPLC trace for SLM169/Derivative 6

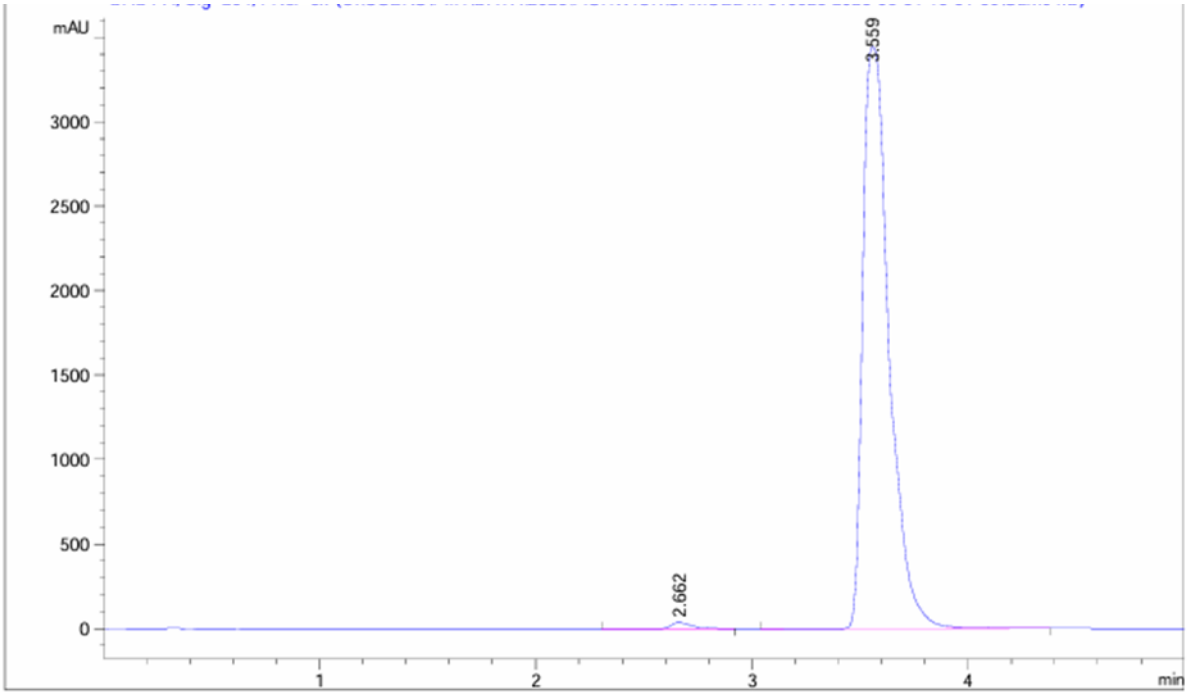


Figure 174; HPLC trace for SLM94/Derivative 7

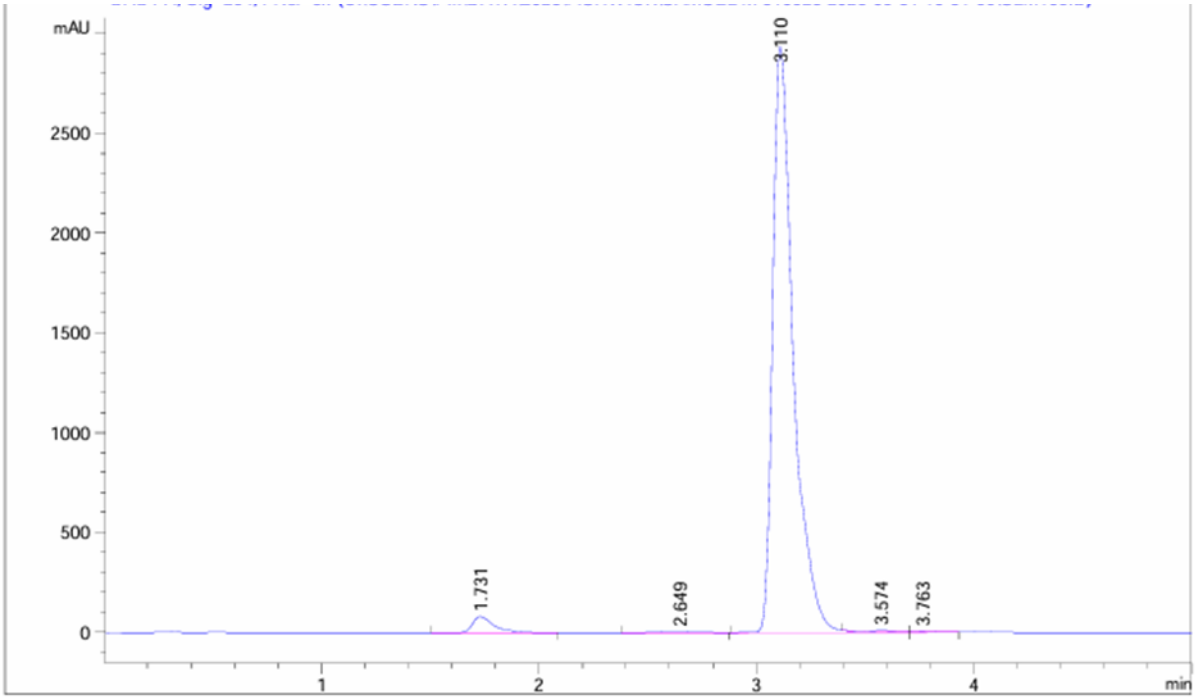


Figure 175; HPLC trace for SLM165/Derivative 8

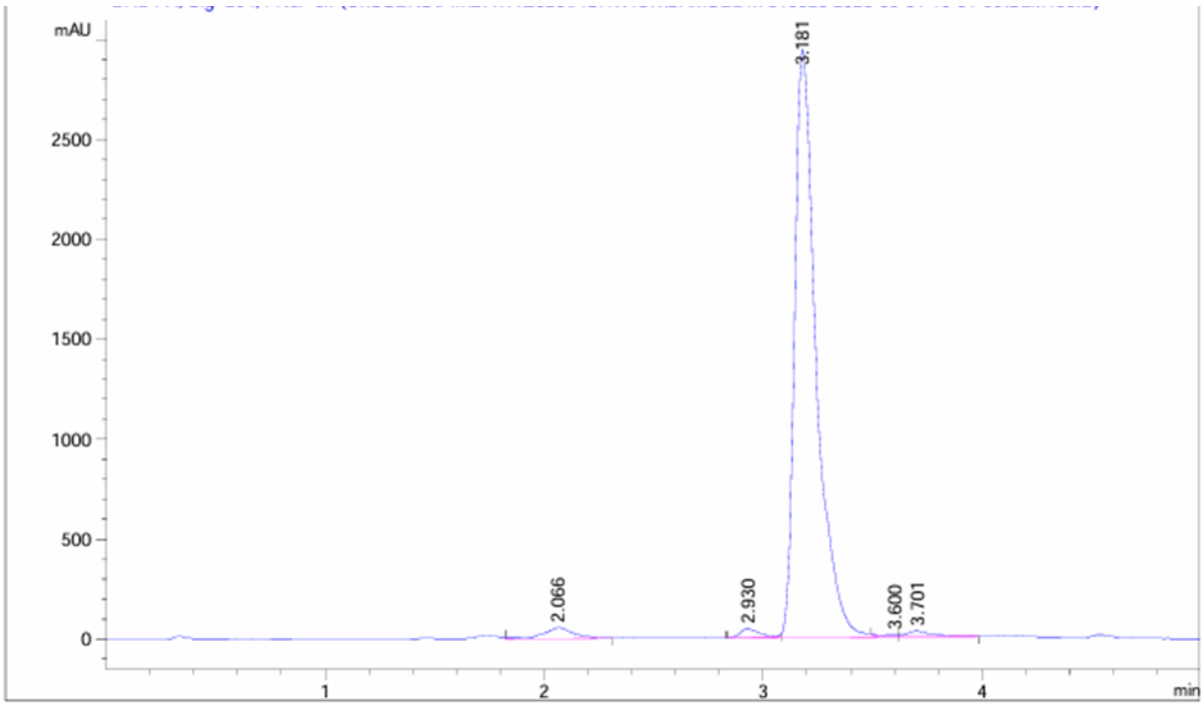


Figure 176; HPLC trace of SLM155/Derivative 9

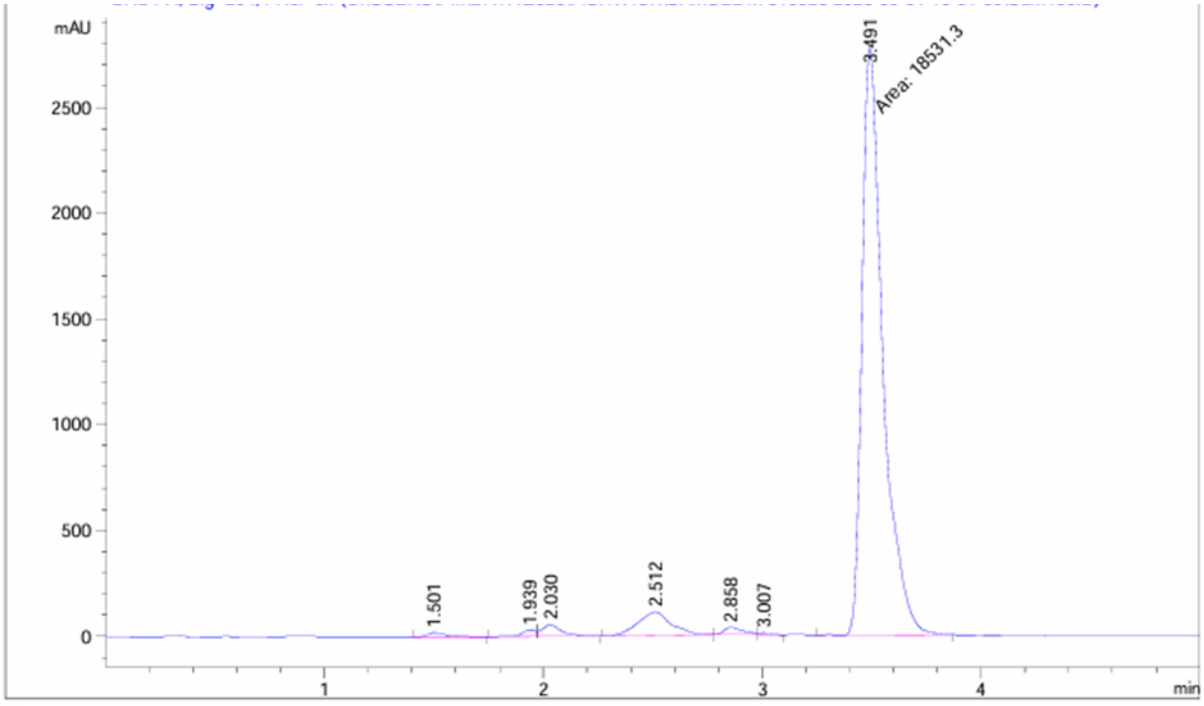


Figure 177; HPLC trace of SLM156/Derivative 10

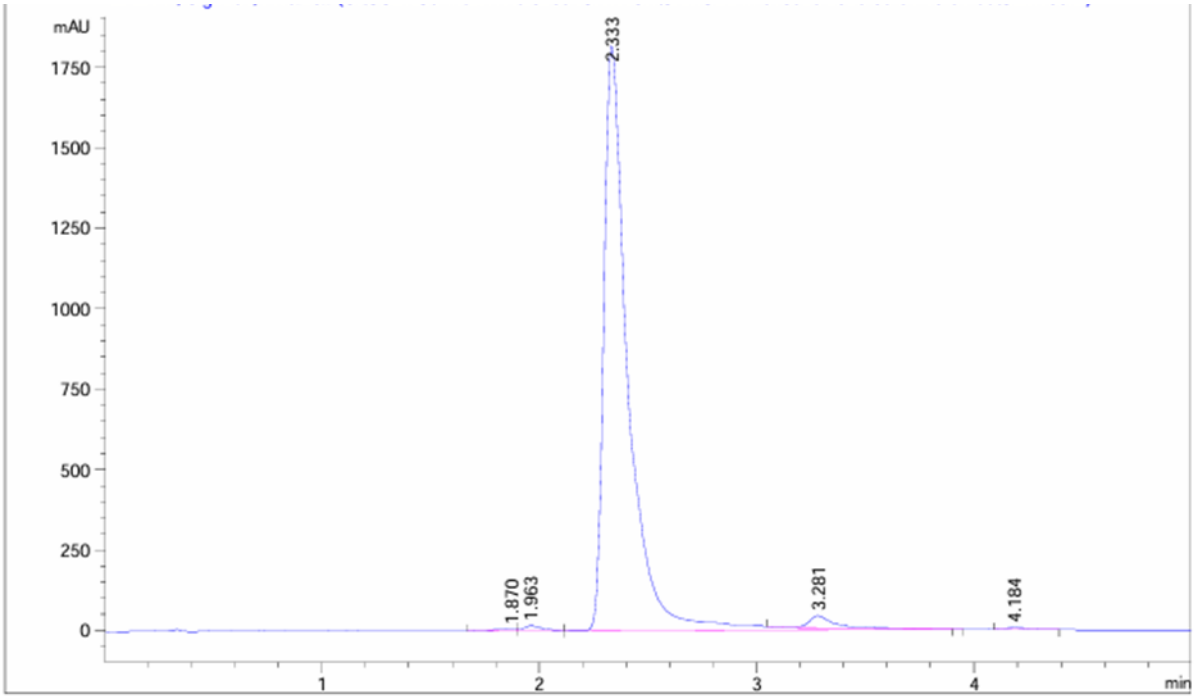


Figure 178; HPLC trace of SLM144/Derivative 11

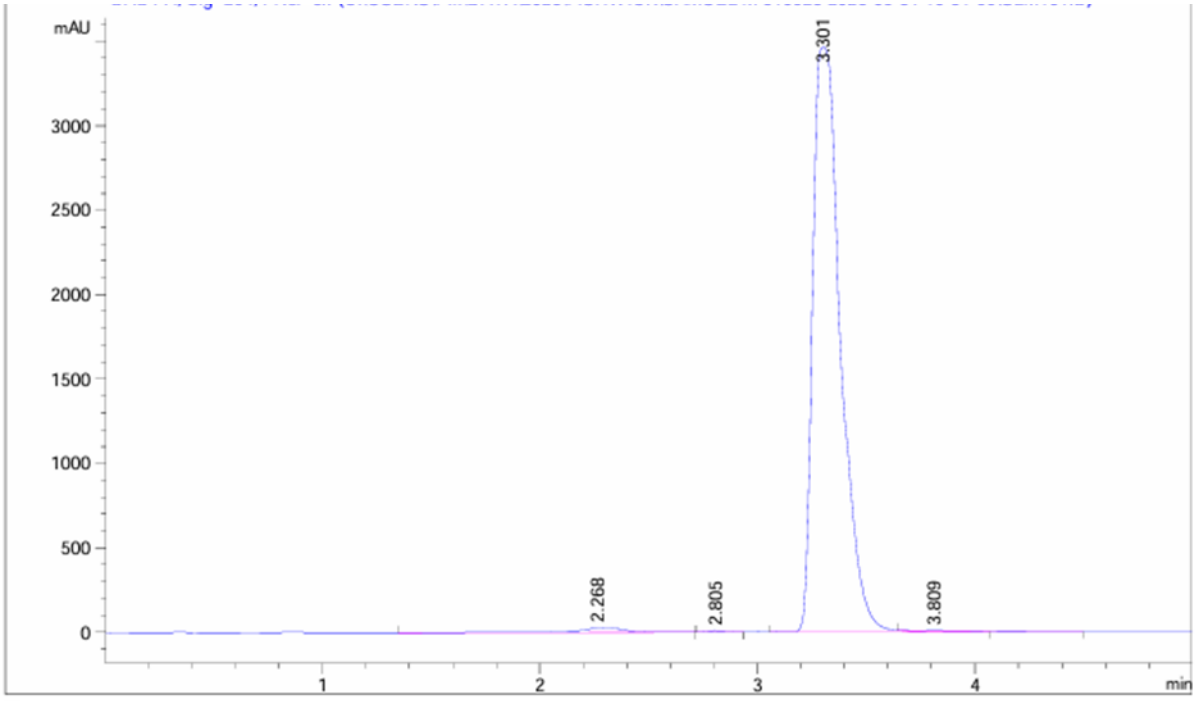


Figure 179; HPLC trace of SLM161/Derivative 12

6.3 High Resolution Mass Spectra relating to Chapter 3

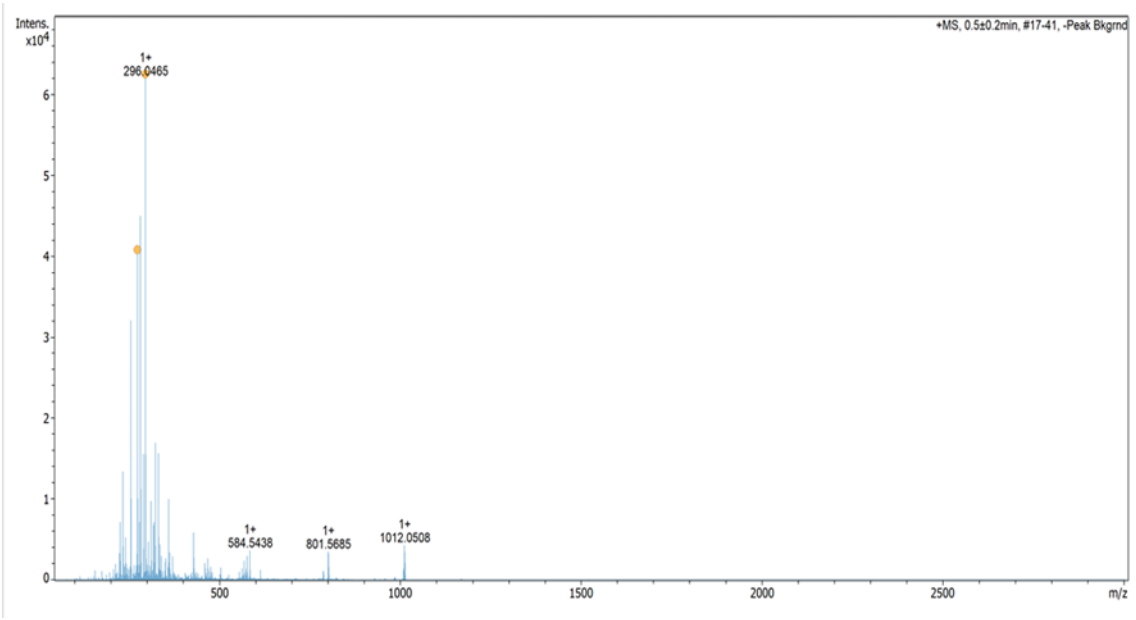


Figure 180; HRMS spectrum of SLM143/Derivative 1

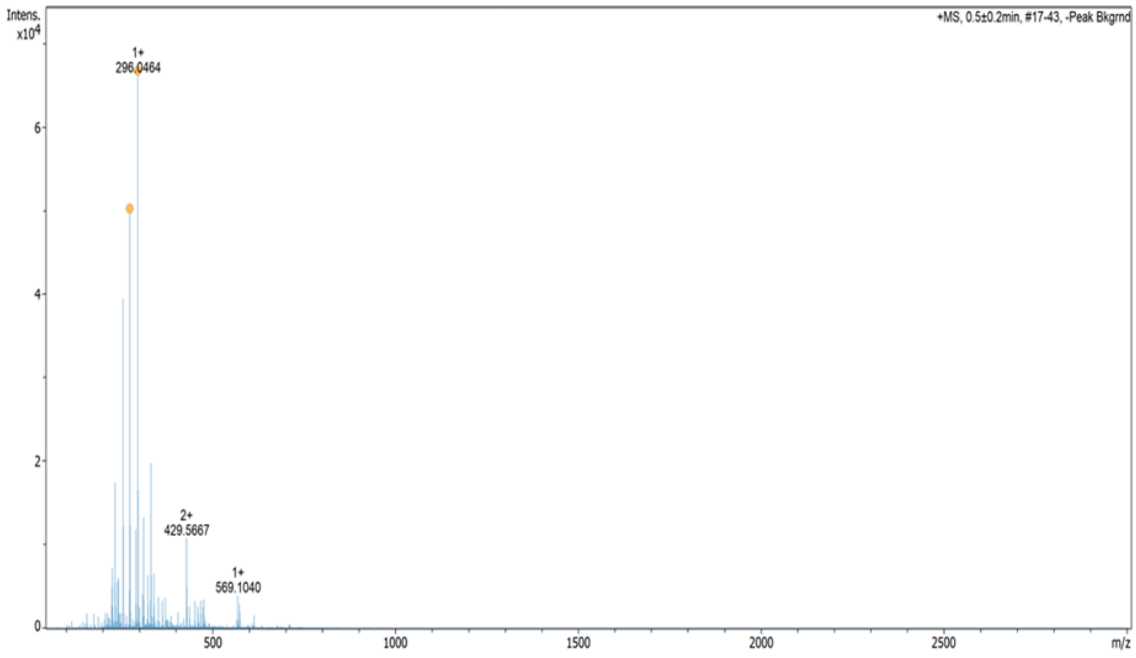


Figure 181; HRMS spectrum of SLM168/Derivative 3

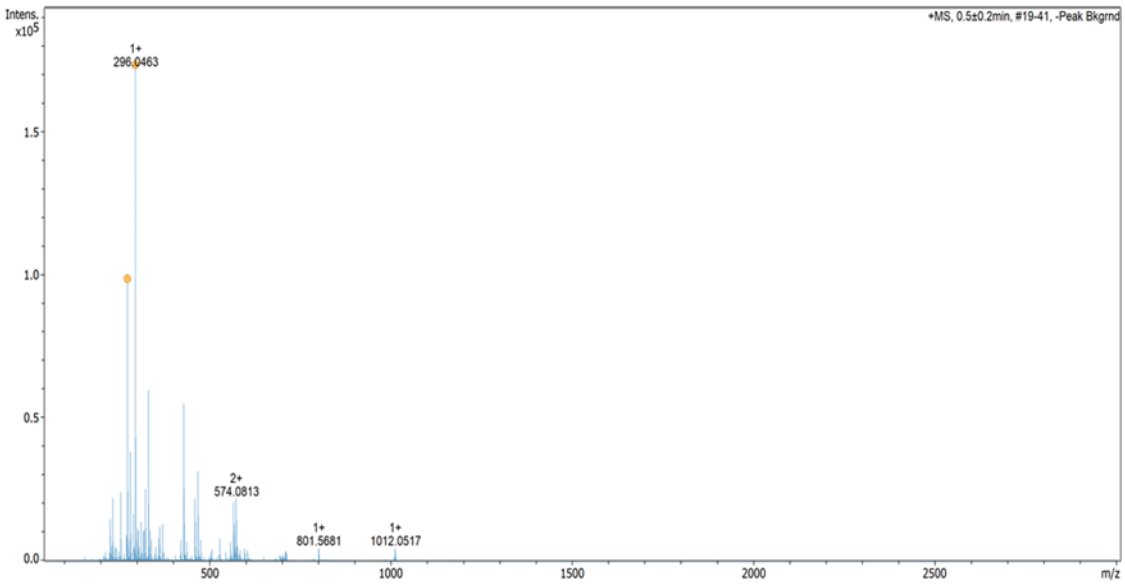


Figure 182; HRMS spectrum of SLM142/Derivative 4

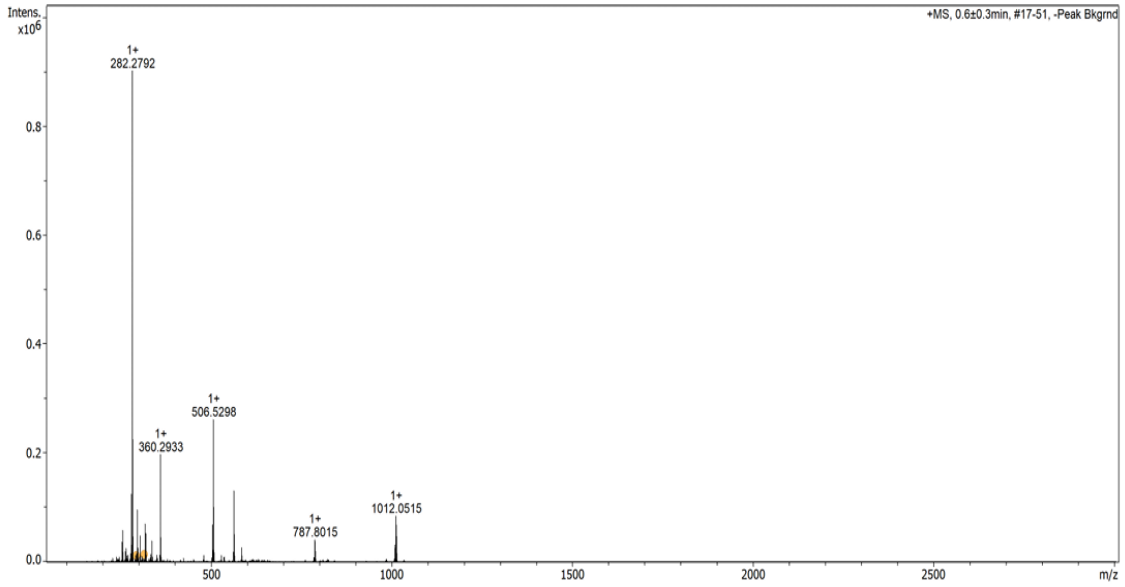


Figure 183; HRMS spectrum of SLM96/Derivative 5

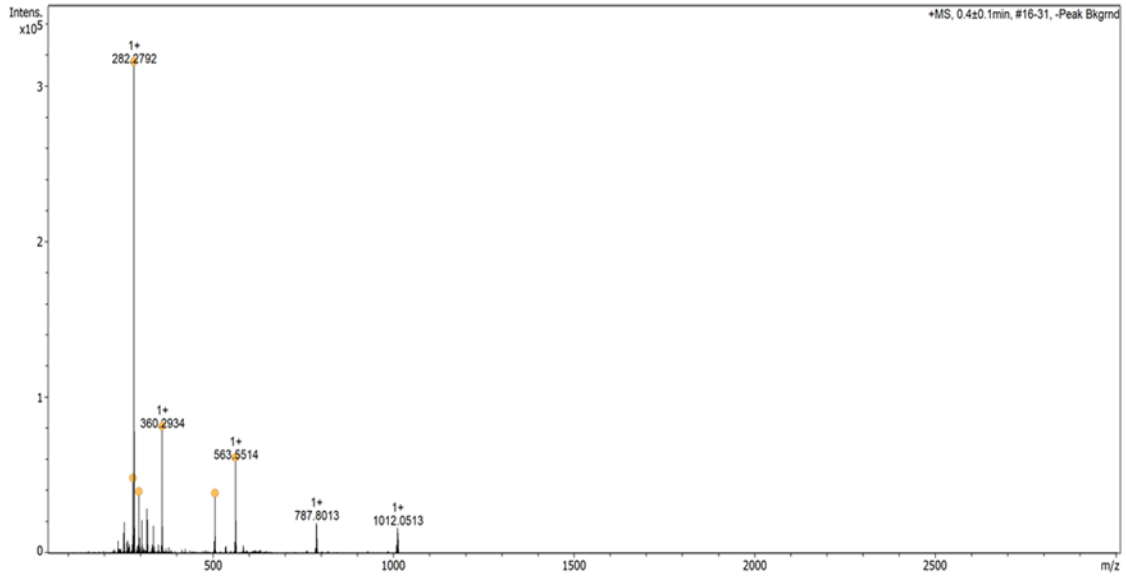


Figure 184; HRMS spectrum of SLM/Derivative 6

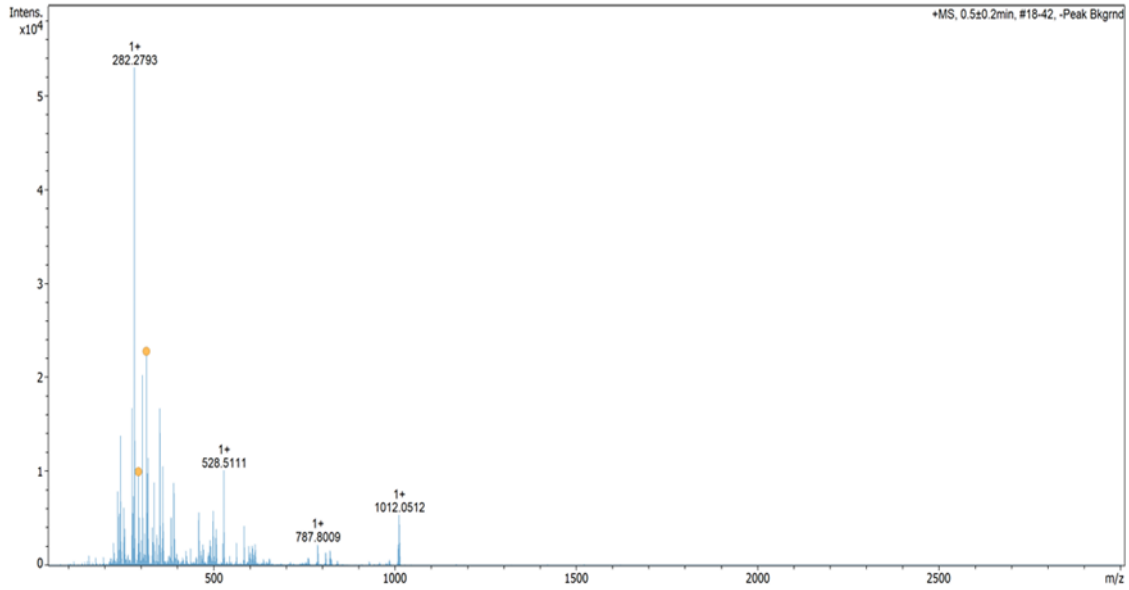


Figure 185; HRMS spectrum of SLM94/Derivative 7

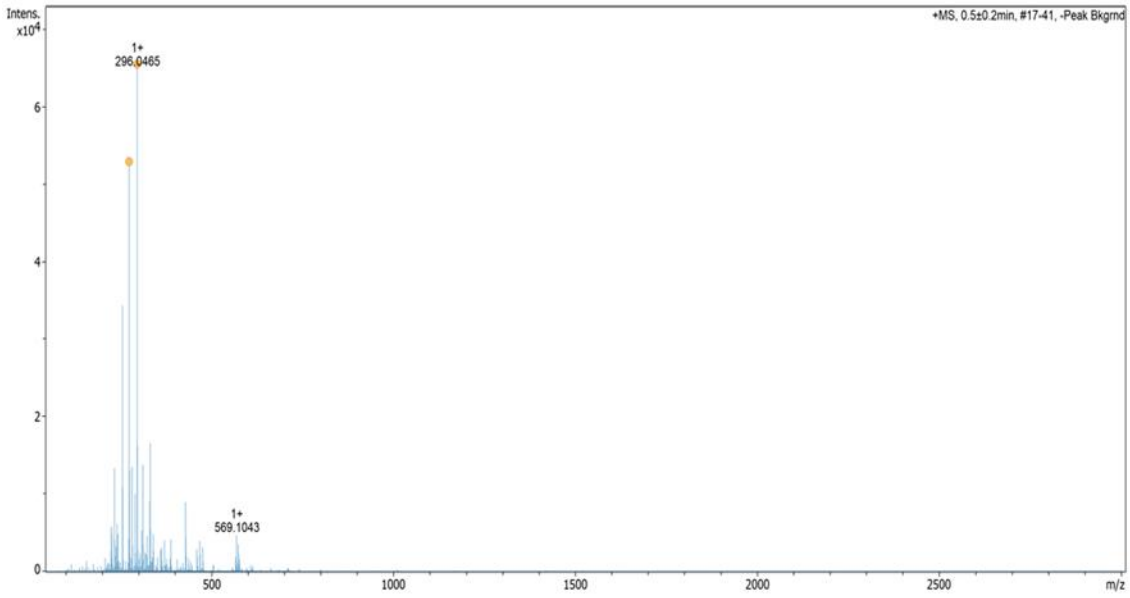


Figure 186; HRMS spectrum of SLM156/Derivative 10

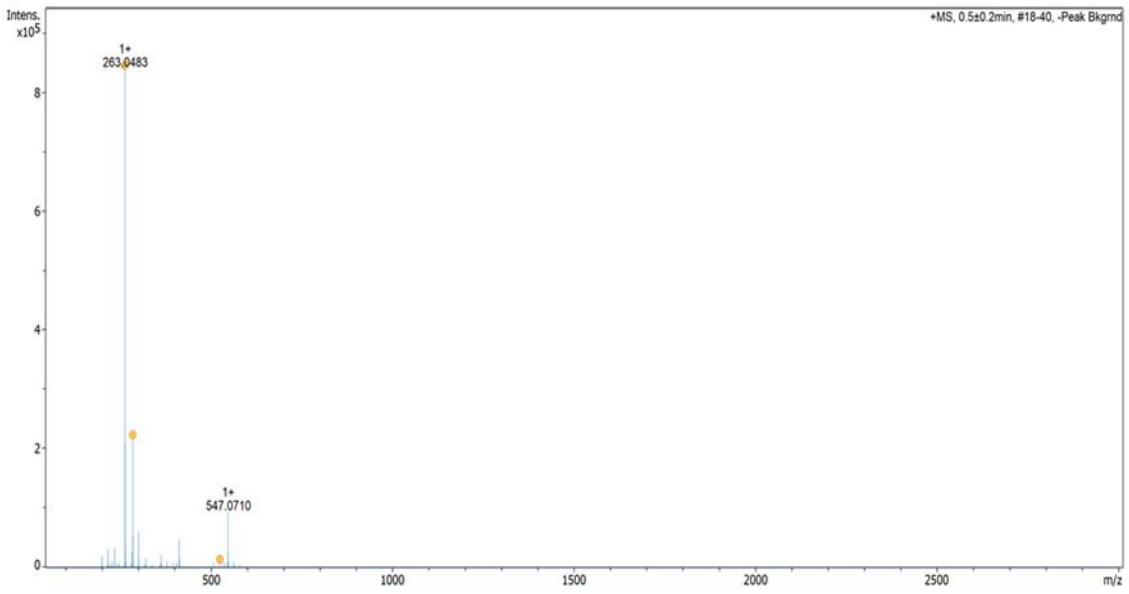


Figure 187; HRMS spectrum of SLM161/Derivative 12

Chapter 7 Bibliography

1. Brog, J.-P., Chanez, C.-L., Crochet, A. & Fromm, K. M. Polymorphism, what it is and how to identify it: a systematic review. *RSC Adv* **3**, 16905 (2013).
2. J. Leibig and F. Wohler. Untersuchungen über das Radikal der Benzoesäure. *Annalen der Pharmacie* 514–514 (1832).
3. Lehmann, O. Ueber physikalische Isomerie. *Z Kristallogr Cryst Mater* **1**, 97–131 (1877).
4. Kofler, L. & Kofler, A. *Thermo-Mikro-Methoden : Zur Kennzeichnung Organischer Stoffe Und Stoffgemische*. (Verlag Chemie, 1954).
5. Krauskopf, K. B. Fusion Methods in Chemical Microscopy Walter C. McCrone, Jr., Interscience, New York, 1957. 307 pp. *Science (1979)* **125**, 1045–1046 (1957).
6. Haleblian, J. & McCrone, W. Pharmaceutical Applications of Polymorphism. *J Pharm Sci* **58**, 911–929 (1969).
7. Macek, T. J. The physical and chemical problems inherent in the formulation of dosage forms for new pharmaceuticals. I. *Am J Pharm Sci Support Public Health* **137**, 217–38 contd (1965).
8. Higuchi, W. I., Bernardo, P. D. & Mehta, S. C. Polymorphism and Drug Availability II. *J Pharm Sci* **56**, 200–207 (1967).
9. Higuchi, W. I., Lau, P. K., Higuchi, T. & Shell, J. W. Polymorphism and Drug Availability. *J Pharm Sci* **52**, 150–153 (1963).
10. Trivedi, J., Shell, J. W. & Biles, J. A. Some Physical and Crystallographic Properties of the Ouabain Hydrates. *Journal of the American Pharmaceutical Association (Scientific ed.)* **48**, 583–587 (1959).
11. Cleverley, B. & Williams, P. P. Polymorphism in substituted barbituric acids. *Tetrahedron* **7**, 277–288 (1959).
12. Higuchi, T. Mechanism of sustained-action medication. Theoretical analysis of rate of release of solid drugs dispersed in solid matrices. *J Pharm Sci* **52**, 1145–1149 (1963).
13. Mesley, R. J., Clements, R. L., Flaherty, B. & Goodhead, K. The polymorphism of phenobarbitone. *Journal of Pharmacy and Pharmacology* **20**, 329–340 (2011).

14. Yang, S. S. & Guillory, J. K. Polymorphism in Sulfonamides. *J Pharm Sci* **61**, 26–40 (1972).

15. Otto, C. E. Fusion methods in chemical microscopy (McCrone, Walter C., Jr.). *J Chem Educ* **34**, A410 (1957).

16. *GLAXO, INC. and Glaxo Group Limited, Plaintiffs, v. NOVOPHARM LTD., Defendant.* (1993). (Accessible at https://law.justia.com/cases/federal/appellate-courts/F3/52/1043/573023/)

17. Hemphill, C. S. Drug Patent Settlements Between Rivals: A Survey. *SSRN Electronic Journal* (2007) doi:10.2139/ssrn.969492.

18. Bauer, J. *et al.* Ritonavir: An Extraordinary Example of Conformational Polymorphism. *Pharm Res* **18**, 859–866 (2001).

19. Chemburkar, S. R. *et al.* Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development. *Org Process Res Dev* **4**, 413–417 (2000).

20. Jarring, K., Larsson, T., Stensland, B. & Ymén, I. Thermodynamic Stability and Crystal Structures for Polymorphs and Solvates of Formoterol Fumarate. *J Pharm Sci* **95**, 1144–1161 (2006).

21. SUN, C. Solid-state properties and crystallization behavior of PHA-739521 polymorphs. *Int J Pharm* **319**, 114–120 (2006).

22. Campeta, A. M. *et al.* Development of a Targeted Polymorph Screening Approach for a Complex Polymorphic and Highly Solvating API. *J Pharm Sci* **99**, 3874–3886 (2010).

23. Pfund, L. Y. & Matzger, A. J. Towards Exhaustive and Automated High-Throughput Screening for Crystalline Polymorphs. *ACS Comb Sci* **16**, 309–313 (2014).

24. Lévesque, A., Maris, T. & Wuest, J. D. ROY Reclaims Its Crown: New Ways to Increase Polymorphic Diversity. *J Am Chem Soc* **142**, 11873–11883 (2020).

25. Li, X. *et al.* The Twelfth Solved Structure of ROY: Single Crystals of Y04 Grown from Melt Microdroplets. *Cryst Growth Des* **20**, 7093–7097 (2020).

26. Bowskill, D. H., Sugden, I. J., Konstantinopoulos, S., Adjiman, C. S. & Pantelides, C. C. Crystal Structure Prediction Methods for Organic Molecules: State of the Art. *Annu Rev Chem Biomol Eng* **12**, 593–623 (2021).

27. Price, S. L., Braun, D. E. & Reutzel-Edens, S. M. Can computed crystal energy landscapes help understand pharmaceutical solids? *Chemical Communications* **52**, 7065–7077 (2016).

28. Braun, D. E., Bhardwaj, R. M., Florence, A. J., Tocher, D. A. & Price, S. L. Complex Polymorphic System of Gallic Acid—Five Monohydrates, Three Anhydrates, and over 20 Solvates. *Cryst Growth Des* **13**, 19–23 (2013).

29. Habermehl, S., Mörschel, P., Eisenbrandt, P., Hammer, S. M. & Schmidt, M. U. Structure determination from powder data without prior indexing, using a similarity measure based on cross-correlation functions. *Acta Crystallogr B Struct Sci Cryst Eng Mater* **70**, 347–359 (2014).

30. Baias, M. *et al. De Novo* Determination of the Crystal Structure of a Large Drug Molecule by Crystal Structure Prediction-Based Powder NMR Crystallography. *J Am Chem Soc* **135**, 17501–17507 (2013).

31. Reilly, A. M. *et al.* Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallogr B Struct Sci Cryst Eng Mater* **72**, 439–459 (2016).

32. Neves, B. J. *et al.* QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Front Pharmacol* **9**, (2018).

33. Rudrapal, M. & Chetia, D. Virtual Screening, Molecular Docking and QSAR Studies in Drug Discovery and Development Programme. *Journal of Drug Delivery and Therapeutics* **10**, 225–233 (2020).

34. Bauer, J. *et al.* Ritonavir: An Extraordinary Example of Conformational Polymorphism. *Pharm Res* **18**, 859–866 (2001).

35. Oganov, A. R. Crystal structure prediction: reflections on present status and challenges. *Faraday Discuss* **211**, 643–660 (2018).

36. Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr B Struct Sci Cryst Eng Mater* **72**, 171–179 (2016).

37. Sharma, P., Sharma, K. & Nandave, M. Computational approaches in drug discovery and design. in *Computational Approaches in Drug Discovery, Development and Systems Pharmacology* 53–93 (Elsevier, 2023). doi:10.1016/B978-0-323-99137-7.00009-5.

38. Galek, P. T. A., Pidcock, E., Wood, P. A., Bruno, I. J. & Groom, C. R. One in half a million: a solid form informatics study of a pharmaceutical crystal structure. *CrystEngComm* **14**, 2391–2403 (2012).

39. Messenheimer, J. A. Lamotrigine. *Epilepsia* **36**, (1995).

40. Weisler, R. H. *et al.* Discovery and development of lamotrigine for bipolar disorder: A story of serendipity, clinical observations, risk taking, and persistence. *J Affect Disord* **108**, 1–9 (2008).

41. Goldsmith, D. R., Wagstaff, A. J., Ibbotson, T. & Perry, C. M. lamotrigine. *Drugs* **63**, 2029–2050 (2003).

42. Sridhar, B. & Ravikumar, K. lamotrigine, an antiepileptic drug, and its chloride and nitrate salts. *Acta Crystallogr C* **65**, o460–o464 (2009).

43. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings 1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3–25. 1. *Adv Drug Deliv Rev* **46**, 3–26 (2001).

44. Allen, F. H. & Taylor, R. Research applications of the Cambridge Structural Database (CSD). *Chem Soc Rev* **33**, 463 (2004).

45. MacRae, C. F. *et al.* Mercury 4.0: From visualization to analysis, design and prediction. *J Appl Crystallogr* **53**, 226–235 (2020).

46. Bruno, I. J. *et al.* Retrieval of Crystallographically-Derived Molecular Geometry Information. *J Chem Inf Comput Sci* **44**, 2133–2144 (2004).

47. Doriguetto, A. C. *et al.* 2,2′,4-Trihydroxybenzophenone: Crystal Structure, and Anti-Inflammatory and Antioxidant Activities. *Chem Biodivers* **4**, 488–499 (2007).

48. Gavezzotti, A. *Molecular Aggregation*. (Oxford University PressOxford, 2006). doi:10.1093/acprof:oso/9780198570806.001.0001.

49. Novikov, A. S. IsoStar Program Suite for Studies of Noncovalent Interactions in Crystals of Chemical Compounds. *Crystals (Basel)* **11**, 162 (2021).

50. Wood, P. A. *et al.* Evaluation of molecular crystal structures using Full Interaction Maps. *CrystEngComm* **15**, 65–72 (2013).

51. Arunan, E. *et al.* Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure and Applied Chemistry* **83**, 1637–1641 (2011).

52. Galek, P. T. A., Fábián, L., Motherwell, W. D. S., Allen, F. H. & Feeder, N. Knowledge-based model of hydrogen-bonding propensity in organic crystals. *Acta Crystallogr B* **63**, 768–782 (2007).

53. Costantino, L. & Barlocco, D. Privileged Structures as Leads in Medicinal Chemistry. *Curr Med Chem* **13**, 65–85 (2006).

54. Yoon, T. P. & Jacobsen, E. N. Privileged Chiral Catalysts. *Science (1979)* **299**, 1691–1693 (2003).

55. Kuhnert-Brandst□tter, M. & Wunsch, S. Polymorphie und Mischkristallbildung bei Sulfonamiden und verwandten Verbindungen. *Mikrochim Acta* **57**, 1297–1307 (1969).

56. Yang, S. S. & Guillory, J. K. Polymorphism in Sulfonamides. *J Pharm Sci* **61**, 26–40 (1972).

57. Hughes, D. S., Hursthouse, M. B., Threlfall, T. & Tavener, S. A new polymorph of sulfathiazole. *Acta Crystallogr C* **55**, 1831–1833 (1999).

58. Zhang, G. G. Z. *et al.* Crystallization and Transitions of Sulfamerazine Polymorphs. *J Pharm Sci* **91**, 1089–1100 (2002).

59. Lutker, K. M. & Matzger, A. J. Crystal polymorphism in a carbamazepine derivative: Oxcarbazepine. *J Pharm Sci* **99**, 794–803 (2010).

60. López-Mejías, V., Kampf, J. W. & Matzger, A. J. Nonamorphism in Flufenamic Acid and a New Record for a Polymorphic Compound with Solved Structures. *J Am Chem Soc* **134**, 9872–9875 (2012).

61. Sacchi, P., Reutzel-Edens, S. M. & Cruz-Cabeza, A. J. The unexpected discovery of the ninth polymorph of tolfenamic acid. *CrystEngComm* **23**, 3636–3647 (2021).

62. Li, H., Stowell, J. G., Borchardt, T. B. & Byrn, S. R. Synthesis, Conformational Polymorphism, and Construction of a *G* − *T* Diagram of 2-[(2-Nitrophenyl)amino]-3-thiophenecarbonitrile. *Cryst Growth Des* **6**, 2469–2474 (2006).

63. He, X., Griesser, U. J., Stowell, J. G., Borchardt, T. B. & Byrn, S. R. Conformational Color Polymorphism and Control of Crystallization of 5-Methyl-2-[(4-methyl-2-nitrophenyl)amino]-3-thiophenecarbonitrile. *J Pharm Sci* **90**, 371–388 (2001).

64. Nogueira, B. A. *et al.* Portrayal of the color polymorphism in the 5-acetyl-derivative of ROY. *CrystEngComm* **24**, 1459–1474 (2022).

65. Nogueira, B. A. *et al.* Color Polymorphs of ROY-ol. *Cryst Growth Des* **22**, 5375–5389 (2022).

66. Klose, C. *et al.* Fenamates as TRP channel blockers: mefenamic acid selectively blocks TRPM3. *Br J Pharmacol* **162**, 1757–1769 (2011).

67. Uzoh, O. G., Cruz-Cabeza, A. J. & Price, S. L. Is the Fenamate Group a Polymorphophore? Contrasting the Crystal Energy Landscapes of Fenamic and Tolfenamic Acids. *Cryst Growth Des* **12**, 4230–4239 (2012).

68. Shevchenko, V.P., Nagaev, I.Y., Kuznetsov, Y.V. *et al.* A Synthesis of Tritium-Labeled Olanzapine. *Russ J Bioorg Chem* **31**, 378–382 (2005).

69. Yu, L. *et al.* Thermochemistry and conformational polymorphism of a hexamorphic crystal system. *J Am Chem Soc* **122**, 585–591 (2000).

70. Chen, S., Guzei, I. A. & Yu, L. New polymorphs of ROY and new record for coexisting polymorphs of solved structures. *J Am Chem Soc* **127**, 9881–9885 (2005).

71. Yu, L. Nucleation of One Polymorph by Another. *J Am Chem Soc* **125**, 6380–6381 (2003).

72. Chen, S., Xi, H. & Yu, L. Cross-nucleation between ROY polymorphs. *J Am Chem Soc* **127**, 17439–17444 (2005).

73. Tan, M. *et al.* ROY revisited, again: the eighth solved structure. *Faraday Discuss* **211**, 477–491 (2018).

74. Gushurst, K. S., Nyman, J. & Boerrigter, S. X. M. The PO13 crystal structure of ROY. *CrystEngComm* **21**, 1363–1368 (2019).

75. Tyler, A. R. *et al.* Encapsulated Nanodroplet Crystallization of Organic-Soluble Small Molecules. *Chem* **6**, 1755–1765 (2020).

76. Cherukuvada, S. & Nangia, A. Eutectics as improved pharmaceutical materials: design, properties and characterization. *Chem. Commun.* **50**, 906–923 (2014).

77. Lutker, K. M., Tolstyka, Z. P. & Matzger, A. J. Investigation of a Privileged Polymorphic Motif: A Dimeric ROY Derivative. *Cryst Growth Des* **8**, 136–139 (2008).

78. Stephenson, G. A. *et al.* Conformational and color polymorphism of 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile. *J Pharm Sci* **84**, 1385–1386 (1995).

79. Li, X. *et al.* The Twelfth Solved Structure of ROY: Single Crystals of Y04 Grown from Melt Microdroplets. *Cryst Growth Des* **20**, 7093–7097 (2020).

80. Bruno, I. J. *et al.* New software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallogr B* **58**, 389–397 (2002).

81. Harty, E. L. *et al.* Reversible piezochromism in a molecular wine-rack. *Chem Comm* **51**, 10608–10611 (2015).

82. Funnell, N. P., Bull, C. L., Ridley, C. J. & Capelli, S. Structural behaviour of OP-ROY at extreme conditions. *CrystEngComm* **21**, 4473–4483 (2019).

83. Macrae, C. F. *et al. Mercury 4.0* : from visualization to analysis, design and prediction. *J Appl Crystallogr* **53**, 226–235 (2020).

84. Allen, F. H., Harris, S. E. & Taylor, R. Comparison of conformer distributions in the crystalline state with conformational energies calculated by ab initio techniques. *J Comput Aided Mol Des* **10**, 247–254 (1996).

85. Beran, G. J. O. *et al.* How many more polymorphs of ROY remain undiscovered. *Chem Sci* **13**, 1288–1297 (2022).

86. Cruz-Cabeza, A. J. & Bernstein, J. Conformational Polymorphism. *Chem Rev* **114**, 2170–2191 (2014).

87. Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr B Struct Sci Cryst Eng Mater* **72**, 171–179 (2016).

88. Macrae, C. F. *et al.* Mercury 4.0: from visualization to analysis, design and prediction. *J Appl Crystallogr* **53**, 226–235 (2020).

89. Price, L. S. *et al.* A molecular picture of the problems in ensuring structural purity of tazofelone. *J Mol Struct* **1078**, 26–42 (2014).

90. Galek, P. T. A., Fábián, L., Motherwell, W. D. S., Allen, F. H. & Feeder, N. Knowledge-based model of hydrogen-bonding propensity in organic crystals. *Acta Crystallogr B* **63**, 768–782 (2007).

91. Vasileiadis, M., Kazantsev, A. V., Karamertzanis, P. G., Adjiman, C. S. & Pantelides, C. C. The polymorphs of ROY: application of a systematic crystal structure prediction technique. *Acta Crystallogr B* **68**, 677–685 (2012).

92. Dossetter, A. G., Griffen, E. J. & Leach, A. G. Matched Molecular Pair Analysis in drug discovery. *Drug Discov Today* **18**, 724–731 (2013).

93. Wood, P. A. *et al.* Evaluation of molecular crystal structures using Full Interaction Maps. *CrystEngComm* **15**, 65–72 (2013).

94. López-Mejías, V. & Matzger, A. J. Structure–Polymorphism Study of Fenamates: Toward Developing an Understanding of the Polymorphophore. *Cryst Growth Des* **15**, 3955–3962 (2015).

95. Leyva-Pérez, A., Cabrero-Antonino, J. R. & Corma, A. Bifunctional solid catalysts for chemoselective hydrogenation–cyclisation–amination cascade reactions of relevance for the synthesis of pharmaceuticals. *Tetrahedron* **66**, 8203–8209 (2010).

96. Gewald, K., Schinke, E. & Böttcher, H. Heterocyclen aus CH-aciden Nitrilen, VIII. 2-Amino-thiophene aus methylenaktiven Nitrilen, Carbonylverbindungen und Schwefel. *Chem Ber* **99**, 94–100 (1966).

97. Puterova, Z., Krutošíková, A. & Végh, D. Gewald reaction: synthesis, properties and applications of substituted 2-aminothiophenes. *Arkivoc* **2010**, 209–246 (2010).

98. Wang, T. *et al.* An efficient one-pot synthesis of substituted 2-aminothiophenes via three-component gewald reaction catalyzed by l -proline. *Synlett* 1351–1354 (2010)

99. Cruz-Cabeza, A. J., Reutzel-Edens, S. M. & Bernstein, J. Facts and fictions about polymorphism. *Chem Soc Rev* **44**, 8619–8635 (2015).

100. Bhandary, S., Gonde, S. & Chopra, D. Dissecting the Conformational and Interaction Topological Landscape of *N* -ethynylphenylbenzamide by the Device of Polymorphic Diversity. *Cryst Growth Des* **19**, 1072–1085 (2019).

101. Leyva-Pérez, A., Cabrero-Antonino, J. R. & Corma, A. Bifunctional solid catalysts for chemoselective hydrogenation–cyclisation–amination cascade reactions of relevance for the synthesis of pharmaceuticals. *Tetrahedron* **66**, 8203–8209 (2010).

102. Wang, T. *et al.* An efficient one-pot synthesis of substituted 2-aminothiophenes via three-component Gewald reaction catalyzed by l -proline. *Synlett* 1351–1354 (2010)