MONASH University

The Synergy of Copolymerization Chemistry and Artificial Intelligence

Emma Van de Reydt

Bachelor of Biomedical Sciences [BSc] *and Master of Biomedical Sciences* [MSc]

A thesis submitted for the degree of *Doctor of Philosophy* at

Monash University in 2023

School of Science, Chemistry

Copyright notice

© E. Van de Reydt (*2023*).

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Abstract

Polymers have a vast number of application possibilities in various fields. The tuneability of their properties makes them especially interesting. When designing a polymer for a specific application, it takes a lot of time, effort and money to find the optimal synthesis method. This work focusses on finding more effective ways to go from the desired application to the final polymer product. There are two different strategies that are explored.

The first strategy focusses on making tuneable solubilizers faster and more effective by making the solubilizer in presence of its target molecule. It was theorised that some sequence sections within a statistical copolymer are more effective to enhance the solubility of their hydrophobic target. Copolymers that were made in presence of their target molecule were made and tested to see if the prevalence of those effective sections, and thus the overall effectiveness, was increased. As proof of principle, Piroxicam was used as a model drug. The copolymers investigated are based on the Eudragit® EPO using butyl methacrylate (BMA), 2-(dimethylamino)ethyl methacrylate (DMAEMA) and methyl methacrylate (MMA) as monomers. It was found that "imprinted" copolymers with a ratio of 1:2:1 BMA:DMAEMA:MMA could solubilise 2.5 $mg \cdot mL^{-1}$ of Piroxicam.

The second strategy explores the use of artificial intelligence to use existing data to predict new entries. The hypothesis was that one could improve older predictive methods, such as the Q-e scheme, via machine learning without the need of extensive computational power. Databases for both the propagation rate coefficient (k_p) of homopolymers and the reactivity ratios (r_1 and r_2) of copolymers were carefully made from literature. Parameters for each monomer were added also from literature or from simple predictive software. The use of a random forest was tested to predict r_1 and r_2. For some entries it performed better than the standard Q-e scheme, but this was not always the case. Generating a large but complete database proved to be challenging as well.

Because the r_1 and r_2 database proved to be challenging, it was theorised that the generation of a k_p-predicting database would be an excellent proof of principle to demonstrate that one can make precise predictions based on correlation. Making use of the Leave One Out Cross Validation (LOOCV), the predicted k_p values were plotted against their experimental values. The most successful prediction (r^2 = 0.993) was obtained using a Ridge regression on a limited dataset (n = 36) containing the molecular weight and a distinction between the type of monomer, inductive effect of the tail group, the effect of H-donors and H-acceptors, the dissociation constants and polarizability as parameters.

A streamlined "pushing of a button" to find the desired polymer is still far away but there is potential to modernise chemistry by designing methods and experiments that create large data outputs and using specific algorithms used for small databases.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature:

Print Name: Emma Van de Reydt

Date: 21/03/2023

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Tanja Junkers, for her unwavering support, guidance, and encouragement throughout my PhD journey. Without her extensive knowledge and expertise in polymer science, this thesis would not have been possible. Tanja, thank you so much for this opportunity, your patience and kindness during the difficult COVID-19 times was invaluable.

Secondly, I want to thank my co-supervisor Professor Andrea Robinson for her optimism and ability to make me have a refreshing view on things. I am also grateful to the members of my thesis committee, Professor Kellie Tuck, Professor Philip Chan, and Professor Rico Tabor, for their valuable feedback and insights. Their constructive criticism and encouragement have been instrumental in shaping the direction of this research. Their encouragement was especially appreciated in the final straw.

I would like to acknowledge the members of the PRD group for their collaboration and support. Thank you, everyone, for the contributions, the advice, the support, the nice atmosphere and making coming to work a pleasure. Especially Noam, my student for many months, whose partnership I enjoyed immensely and has been invaluable for this work.

I am also grateful to the Faculty of Science Dean's Postgraduate Research Scholarship (DPRS) and Faculty of Science Dean's International Postgraduate Research Scholarship (DIPRS) for their financial support, which enabled me to pursue my research goals and attend scientific conferences. I am especially grateful for the extension that allowed me to complete this work.

Finally, I would like to thank my family and friends for their unwavering support and encouragement. Their love, patience, and understanding have been a constant source of motivation and inspiration throughout my PhD journey. Becky and Mitchell, Charlotte and Jun, without your friendships I would not have survived lockdown. Thank you for making living on the other side of the world a fantastic experience. I am forever grateful I met you guys. (Especially for introducing Critical Role, watching this together definitely kept me sane during the endless lockdowns).

In addition, I would like to thank my partner Jeroen for his unwavering support, patience, and encouragement. His love and understanding have been a constant source of motivation and inspiration throughout my PhD journey. Jeroen, being separated from you for almost two years was difficult but even from a distance you always believed in us. Your eye for detail helped me elevate many parts of this thesis.

Thank you all for your contributions, support, and encouragement.

Emma

Content

Chapter 1: The Synergy of Copolymerization Chemistry and Artificial Intelligence

1. Solubility of novel pharmaceutical drugs

When developing cures for diseases, it's not always necessary to invent new drugs. In 2020, the cost to develop and market a new medication was estimated to be 1.1 billion USD.[1] On top of that, a large proportion of all new chemical entities are hydrophobic and have limited solubility in polar solvents.[2] Since most modern drugs in development are hydrophobic in nature, the drug delivery inside the human body is more challenging.[3, 4] Attempting to mix hydrophobic substances in water-like media can cause issues like precipitation, phase separation or the formation of a suspension.[2] There are many drug delivery systems, such as microencapsulation, nanoparticles and hydrogels to name a few, that aid in the effectiveness of drug therapy.[5] A well-chosen drug delivery system, improves the application possibilities. Finding new, easy and reproducible methods to create dedicated solubilizers is an interesting challenge.

Defining solubility and permeability

The ease of getting a drug to the market is affected by solubility and permeability, terms that are key parts of the Biopharmaceutics Classification System (BCS). The BCS is comprised of four classes, defined by high or low solubility/permeability (Figure 1-1a). The United States Pharmacopoeia (USP) and British Pharmacopoeia (BP) have both described solubility in parts of solvent required per part of solute. A molecule is considered very soluble when less than one part of solvent is needed and practically insoluble at 10,000 parts or more.[2] Rautio et al. illustrates this classification by plotting the permeability versus the volume required to dissolve the highest dose (Figure 1-1a).[6]

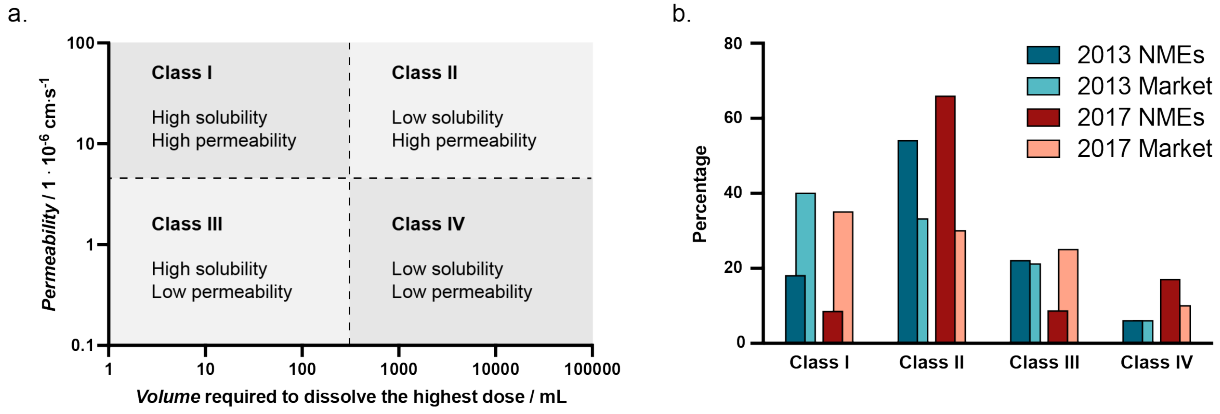


Figure 1-1: (a) Scheme of the different classes in the Biopharmaceutics Classification System (BCS) which are defined by their high or low solubility/permeability. Adapted from Rautio et al.[6] (b) the percentage of market drug molecules according to the Biopharmaceutics Classification System (BCS). For each class, the percentage of New Molecular Entities (NMEs) (dark blue and red) are compares with their market percentage (light blue and pink) for two different years. The values are adapted from Benet et al. (2013)[3] and Nikolakakis et al. (2017)[4]

Solubility is influenced by the polarity of both the solute and the solvent. When molecules are polar, meaning that they have areas with a partial charge, they dissolve easily in polar solvents such as water. To describe a molecules' ability to dissolve in water, the terms hydrophobic and hydrophilic,

water-fearing and water-loving respectively, are often used. Permeability describes how well a molecule can pass through the necessary membranes. Both Benet et al. and Nikolakakis et al. have reported the percentage of drug molecules on the market and the percentage of New Molecular Entities (NME) in 2013[3] and 2017[4] respectively (Figure 1-1b). Although most NME are BCS Class II, only around 30% of the currently marketed drugs are in this class. This disparity between discovered drugs and marketed drugs highlights the value of addressing solubility when developing treatments.

In order for a drug to work as intended, it needs to reach the desired systemic levels and have the correct pharmacological actions. Several methods towards increasing drug solubility in water have been suggested.[2] In 2016, several media outlets highlighted a new liquidized form of aspirin that could pass the blood brain-barrier. The component, IP187B, is a potential candidate towards therapy for brain cancer since research suggests that it could be highly effective against gliobastoma, one of the deadliest forms of brain cancer.[7, 8] Thus, even though aspirin is widely spread and used, it is still involved in significant research.

Solubility enhancement techniques

Solubilization techniques can be classified in three categories: physical modifications, chemical modifications and miscellaneous modifications. One of the most straightforward, efficient and reproducible solubility enhancement techniques is particle size reduction, a physical modification. By increasing the surface-to-volume ratio, there is a greater interaction with the solvent which increases its solubility. However, conventional methods to reduce the particle size rely on mechanical and/or thermal stress which can induce degradation.[2] Similarly, nanosuspensions have an increased surface-to-volume ratio. They can be prepared with precipitation, media milling and/or high-pressure homogenization.[2] Other concepts are to modify the crystal habit (e.g. cocrystallization) or disperse the drug in carriers such as a hydrophilic matrix (e.g. solid dispersions).[9] Changing the pH, the use of a prodrug or complexation are all examples how to enhance solubility via chemical modification.[9]

Miscellaneous modification strategies use a mediator between the hydrophobic drug and the hydrophilic solvent. This mediator can be straightforward, like a cosolvent, or more complex like a surfactant. A surfactant molecule has an amphiphilic nature, meaning it has both a hydrophobic part and a hydrophilic part.[10] The non-polar or hydrophobic part can form strong interactions with hydrophobic drugs whereas the polar part has a strong affinity with the polar solvent. Because of its amphiphilic nature, the surfactant molecules can self-assemble into structures like micelles.[10] An expansion of surfactants is the use of synthetic, amphiphilic block-copolymers. The optimization of controlled polymerization has allowed the yield of precise synthetic polymers with tuneable characteristics. This opens up the possibility to create various shapes such as spheres, ellipsoids or vesicles with self-assembly.[10] Buckinx et al. recently introduced a top-down morphology control in

micellar self-assembly using continuous flow.[11] However, most micelles function as an extern carrier that traps the desired drug and transports it to the desired place of action. These carriers can be precisely tuned to release their contents using antibodies, temperature or pH. Nevertheless, drug solubilizers can be made more specific, and thus more effective, when specific interactions between the drug and the solubilizer are considered as well. A great source for functional groups that induce these interactions is, as in many cases, nature itself: peptides.[12] As stated by Tesauro et al., peptide-based aggregates have many advantages including good biocompatibility and high drug loading capacities.[12] Lawatscheck et al. synthesised a tailored peptide-polyethylene glycol(PEG) conjugate that functions as a specific solubilizer for the hydrophobic drug *N*-phenylamine B4A1 (Figure 1-2).[13] The defined sequence of the peptide part provides a specific, noncovalent drug binding via the peptide/drug interactions. The hydrophilic PEG block ensures that the solubilizer/B4A1 complexes are soluble in water.

However, a disadvantage of natural peptides is that they are sensitive to denaturation which causes them to lose their functionality. Maron et al. uses learning strategies from peptides to synthesize functional precision polymer sequences, hence making the synthetic equivalent to the natural peptide.[14] Moreover, synthetic polymer solubilizers are not limited to the building blocks of natural peptide solubilizers.

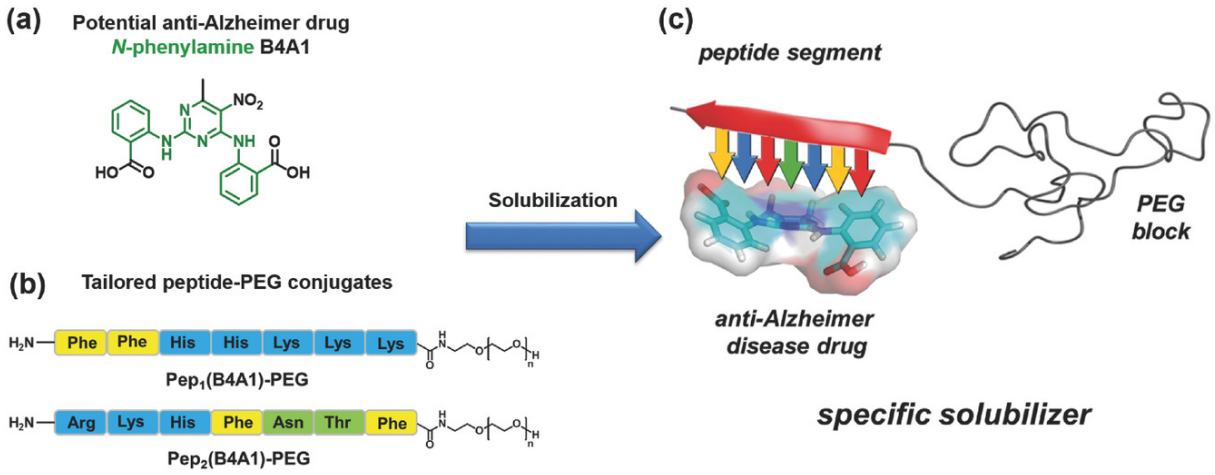


Figure 1-2: The structure of the hydrophobic drug N-phenylamine B4A1 (a) and the structure of the tailored peptide-PEG conjugates (b) Pep_1(B4A1)-PEG and Pep_2(B4A1)-PEG. The functionalities of the peptide segment interact with the B4A1, generating a water-soluble drug conjugate complex due to the water solubility of the polymer part. The abbreviations are as followed: Arg = Arginine, Asn = Asparagine, His = Histamine, Lys = Lysine, Phe = Phenylalanine, Thr = Threonine, PEG = Polyethylene glycol. Insert figure reproduced with permission from Ref[13]. Copyright © Macromolecular Bioscience.

Kano *et al*. uses 2-Methacryloyloxyethyl Phosphorylcholine (MPC) and *n*-Butyl Methacrylate (BMA) to form a copolymer (poly(MPC-co-BMA)) that increases the drug solubility and absorption of miconazole (class II), vidarabine (class III) and griseofulvin (class IV).[15] Poly(MPC-co-BMA) is commercially used as a solubilizer. The MPC molecule is hydrophilic and makes the polymer water soluble. The BMA has a hydrophobic interaction with the hydrophobic drug.[15] The mechanism how the poly(MPC-co-BMA) enhances the oral drug absorption was not clarified. The solubility of

phenytoin was enhanced with a copolymer using 1-vinyl-2-pyrrolidone (VP) and poly(N-isopropylacrylamide-co-vinylpyrrolidone) (NIPAAM) by Widanapathirana *et al.* [16] In here, the NIPAAm was the key for an optimal drug-polymer interaction whereas VP provided aqueous solubility. In conclusion, there are numerous strategies to create polymers, whether it be random copolymers or sequence defined polymers. In order to fully understand the strategy to create the right polymer, these techniques are explained in **section 2**.

2. Polymer chemistry

A polymer is a macromolecule that consists of (relatively) small molecule units bonded together to form a larger molecular structure.[17] The definition of a polymer is very general and can range from proteins, wood and silk, to plastic, rubber and resin. While synthetic polymers are more chemically diverse than biopolymers, biocatalysts like enzymes offer a greater molecular precision.[18] A control of the sequence of the monomer units in a polymer leads to a higher predictability of the polymers properties. Over the years, several techniques were developed to generate sequence-controlled and sequence-defined polymers including (but not limited to) so called “living” free radical, ring-opening and anionic polymerizations. Before elaborating on these techniques, it is important to understand the kinetics behind polymerizations first.

Kinetics

The two main polymerization mechanisms are step-growth and chain-growth polymerization.[19] Step-growth makes use of bi- or multifunctional monomers which are combined together to form the polymer chain. First, monomers will react with each other to form dimers. These dimers will react with any molecular species, form oligomers and eventually a polymer chain (Figure 1-3). On the other hand, chain-growth polymerization makes use of unsaturated monomers which are activated with the help of an initiator. Other unsaturated monomers are joined to the growing polymer chain one at the time (Figure 1-3). Both step-growth and chain-growth polymerization generally lead to non-uniform or polydisperse samples. Optimized polymerization methods are needed to generate precision polymers with precisely controlled molecular structures. Living polymerization is a subclass of chain-growth polymerization. The relationship between the number average molecular weight (M_n) and the conversion is linear (Figure 1-3).[20, 21] There are several techniques that led to the development of precision polymers as illustrated by Lutz et al.[20] In this section, free radical polymerization, which is a part of chain-growth polymerization, will be discussed more in dept.

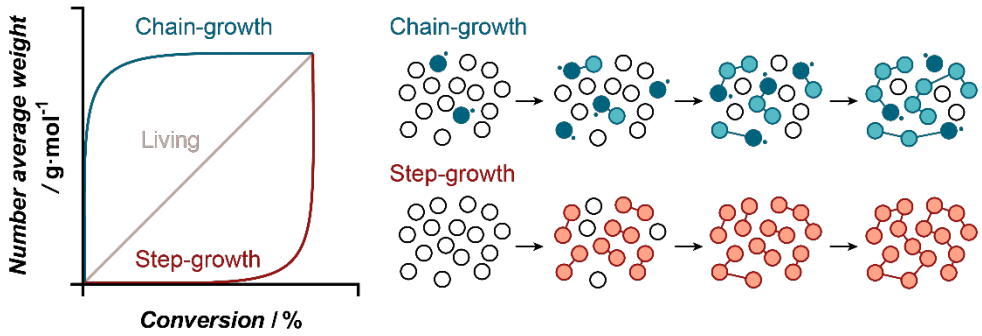


Figure 1-3: The number average molecular weight (M_n) versus conversion plot for chain-growth, living and step-growth polymerization in blue, green and red respectively. The difference between chain- and step-growth polymerization is schematically represented next to the plot.

Free radical polymerization consists of three main reaction steps: initiation, propagation and termination (

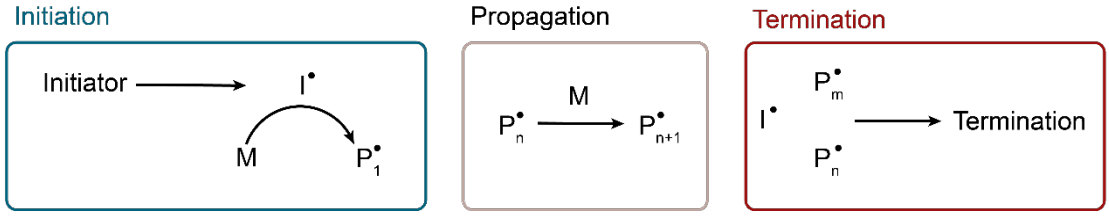


Figure 1-4: Scheme of free radical polymerization divided into three steps: initiation (green), propagation (blue) and termination (red). The I˙ represents an activated initiator, the M a non-active monomer and the P_1˙/P_n˙/P_m˙ a polymer with 1, n and m insertions.

). In the first step, activated monomers with propagating radicals are formed when free radicals from a decomposed initiator react with a monomer. These propagating radicals will form a growing polymer chain. The growth stops during termination where two free radicals form a stable nonradical adduct. It can be noted that the most straightforward free radical polymerization only needs a mixture of monomer and initiator.[21] If two or more monomer species are used in the polymerization, statistical copolymers are formed. These can be completely random, have alternating structures, contain blocky structures or follow other statistical sequences and depend on the reactivity ratio of the comonomers. Even though it is difficult to form defined polymers due to the statistical nature of radical chain-growth polymerization, the sequence of the end product can still be influenced by considering the relative quantity of the monomers' feed and the reactivity ratio. Between two monomers (M_1 and M_2), four different reactions can occur: an activated M_1 or activated M_2 can react with another M_1 or M_2. These reactions each have their own reaction constant (k_{11}, k_{12}, k_{21} and k_{22}). The reactivity ratio (r) of a propagating chain is the ratio of the rate constant for the addition of the same monomer and the rate constant for the addition of the other monomer ($r_1 = k_{11}/k_{12}$ and $r_2 = k_{22}/k_{21}$).[22]

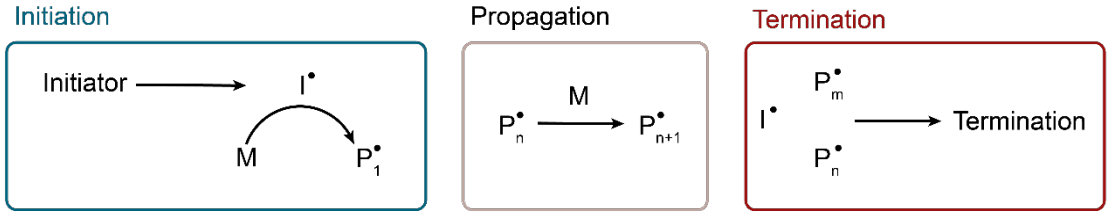


Figure 1-4: Scheme of free radical polymerization divided into three steps: initiation (green), propagation (blue) and termination (red). The I˙ represents an activated initiator, the M a non-active monomer and the P_1˙/P_n˙/P_m˙ a polymer with 1, n and m insertions.

The reaction constant (k_p), or propagation rate coefficient, has a molecular kinetic interpretation often referred to as the Arrhenius (Equation 1).[23, 24]

$$k_p = Ae^{\frac{-E_a}{RT}}$$ Equation 1

Thus, k_p can be written using the pre-exponential frequency factor (A), the activation energy (E_a), the absolute temperature (T) and the gas constant ($R = 8.3145\ J \cdot mol^{-1} \cdot K^{-1}$). The Arrhenius plot shows the relationship between k_p (or $\ln(k_p)$) and the absolute temperature of polymerization where the A and E_a can be extrapolated from the intersection and slope of the curse respectively (Figure 1-5).

The standard technique to determine radical polymerization characteristics is Pulsed Laser Polymerization (PLP). In combination with Size Exclusion Chromatography (SEC), PLP manages to accurately determine k_p when measuring at high frequencies (100-500 Hz). Consequently, in order to determine accurate benchmark k_p values, high frequency lasers are needed, a technology that has only existed for a few decades.[25]

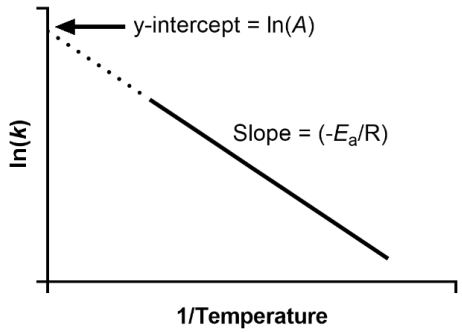


Figure 1-5: Arrhenius plot with the natural logarithm of the reaction constant (ln k) on the Y-axis and a derivative of the absolute temperature (1/T) on the X-axis. The slope and intercept are related to the pre-exponential factor (A) and the activation energy (E_a)

When a polymer consists of one type of monomer, it is often referred to as a homopolymer. Copolymers have multiple building blocks and their final design depends on a variety of factors (Figure 1-6). The inherent property of the monomers is one such factor and the corresponding reactivity ratios (r_1 and r_2) will have an influence on the structure of a polymer.

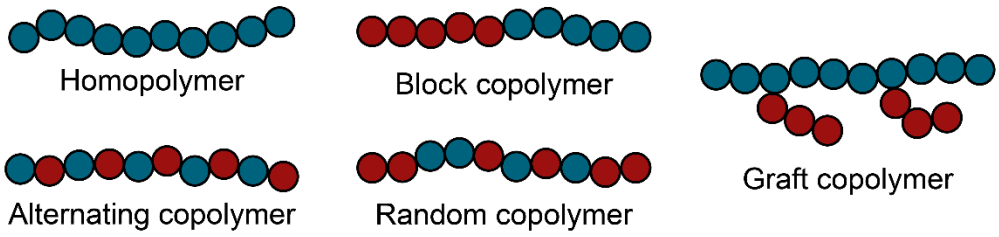


Figure 1-6: Examples of different copolymer designs compared to a linear homopolymer

A large r_1 indicates a propensity for M_1* to add to M_1 whereas a small r_1 means M_1* will add to M_2. Thus, knowledge about the r_1 and r_2 can indicate if the final structure is a homopolymer ($r_1 \approx r_2 >> 1$), block copolymer ($r_1 \approx r_2 > 1$), random ($r_1 \approx r_2 \approx 1$), or alternating ($r_1 \approx r_2 \approx 0$).

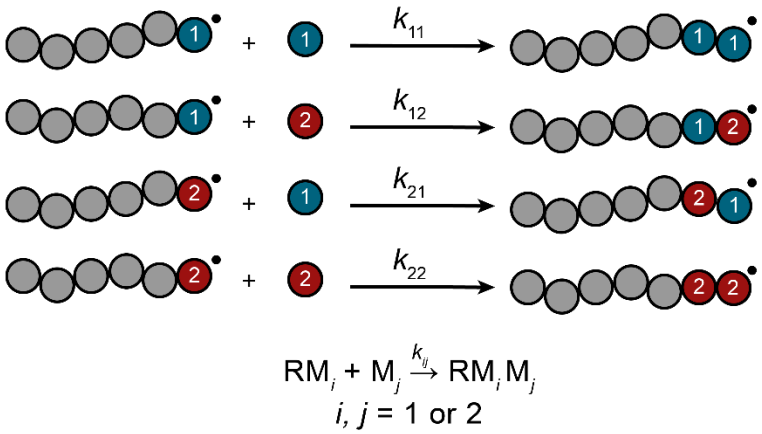


Figure 1-7: Visualisation of the Terminal Model

There are multiple ways to describe copolymerization reactivity ratios. The one that is most straightforward, and often used because of its simplicity, is the Terminal Model described by Mayo et al. (Figure 1-7).[26]

For radical polymerizations, the result of the terminal model would be described by the copolymer equation using the monomer concentrations ([M_1] and [M_2]) and the reactivity ratios (r_1 and r_2) (Equation 2). It is also known as the Mayo-Lewis equation.

$$\frac{d[M_1]}{d[M_2]} = \frac{[M_1](r_1[M_1] + [M_2])}{[M_2]([M_1] + r_2[M_2])} \qquad \text{Equation 2}$$

Alternatively, one can rewrite the equation in terms of mole fractions of monomer in the feed (f_1 and f_2) and mole fractions of the imbedded monomer in the polymer (F_1 and F_2) (Equation 3).

$$F_1 = 1 - F_2 = \frac{r_1 f_1^2 + f_1 f_2}{r_1 f_1^2 + 2 f_1 f_2 + r_2 f_2^2} \qquad \text{Equation 3}$$

Using these equations, it is relatively easy to determine r_1 and r_2. However, with its simplicity come some limitations as well because one has to assume a steady-state system. Thus, one cannot predict low molecular weight polymers.

The penultimate model is a succession of the Terminal Model and takes the effects of the penultimate monomer unit of a growing chain into account. Unlike the Terminal Model, it can describe the rate of polymerization and the composition simultaneously. However, the increase of model parameters results in an increased complexity, and is because of this less popular then the Terminal Model.[26, 27]

The Q-e scheme is a simple extension of the Mayo-Lewis model.[8] The parameters P and Q are the general reactivity of radical i and monomer j, and e is proportional to residual charges in the reacting groups (Equation 4). Substituting the four possible outcomes when a polymer undergoes propagation, we can isolate the reactivity ratios (Equation 5).

$$k_{ij} = P_i Q_j \exp(-e_i e_j) \quad \text{Equation 4}$$

$$r_1 = \frac{k_{11}}{k_{12}} = \frac{Q_1}{Q_2} \exp[-e_1(e_1 - e_2)] \quad \text{Equation 5}$$

Again, the Q-e scheme is limited in its efficiency, especially when introducing new monomers to the system. On top of that, P and Q are dependent on the radical to which the monomer is being bound to.

Controlled polymerizations

Reversible-Deactivation Radical Polymerization (RDRP), formally known as Controlled Radical Polymerization (CRP), allows for more control over the precise molecular weight, dispersity and specific composition. In RDRP, the lifetime of the growing polymer chains is increased by adding a specific compound on top of/instead of the initiator and the monomer. The three most important methods of RDRP are Nitroxide-Mediated Polymerization (NMP)[28], Atom Transfer Radical Polymerization (ATRP)[29] and Reversible Addition-Fragmentation chain-Transfer (RAFT)[30] polymerization (Figure 1-8). NMP and ATRP are based on the principle of reversible deactivation: most radicals are "trapped" in a dormant state which decreases the total radical concentration of the reaction and thus decreases the termination rate. RAFT on the other hand works with a degenerative chain transfer where the termination is decreased by a control agent. Thus, on top of the linear increase of M_n versus conversion, RDRP reactions have a low polydisperse and high end-group fidelity.

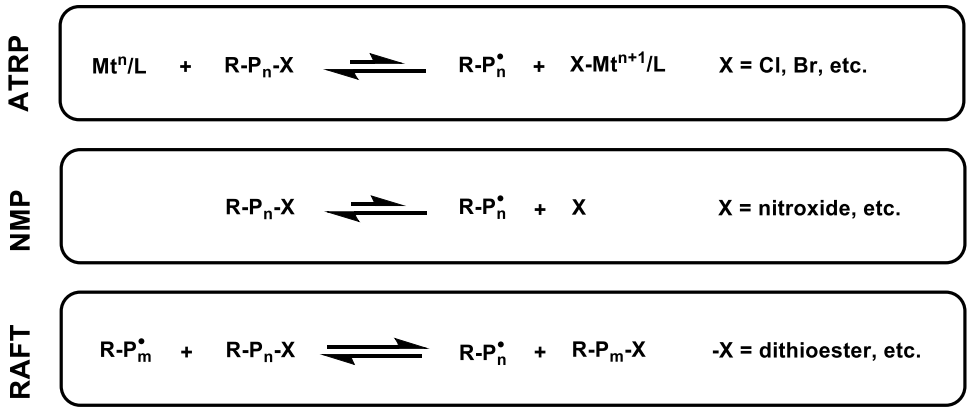


Figure 1-8: Schematic overview of the three main methods for reversible-deactivation radical polymerization (RDRP): nitroxide-mediated polymerization (NMP), atom transfer radical polymerization (ATRP) and reversible addition-fragmentation transfer (RAFT) polymerization. Mt^n/L represents a metal catalyst with a ligand. The -X group is depended on the type of RDRP, the R-groups is a rest group. The P_n is a polymer with n insertions.

RAFT is a versatile technique that is compatible with a wide variety of monomers including styrenes, acrylates, methacrylates, vinyl esters and acrylamides. On top of the initiator and the monomer, a third essential compound is needed in classical RAFT polymerization: the Chain Transfer Agent (CTA) or RAFT-agent. CTA's are typically formed by a dithioester together with a free radical leaving

group (R) and a stabilizing group (Z). There are several components that determine the effectiveness of the RAFT agent: the R-group, the Z-group and the compatibility between the RAFT agent and the monomer. The mechanism of RAFT can be divided into five steps. The first step, the initiation, is the same as the first step of free radical polymerization. The propagating radicals can transfer onto the thiocarbonylthio compound of the RAFT-agent leading to a (possible) detachment of the R-group. New propagating radicals are created in the reinitiation step by the formative radical of the R-group. When the R-group is released from the RAFT-agent, a new active centre is generated (re-initiation step). Thus, the propagating radicals can again transfer onto the thiocarbonylthio compound leading to a dormant intermediate. The main equilibrium is formed between the RAFT adduct radical intermediate and the active propagating radicals, essentially avoiding that two propagating polymers will terminate (Figure 1-9). Since the chains grown simultaneously, the dispersity should be low because all the polymers will have approximately the same length.

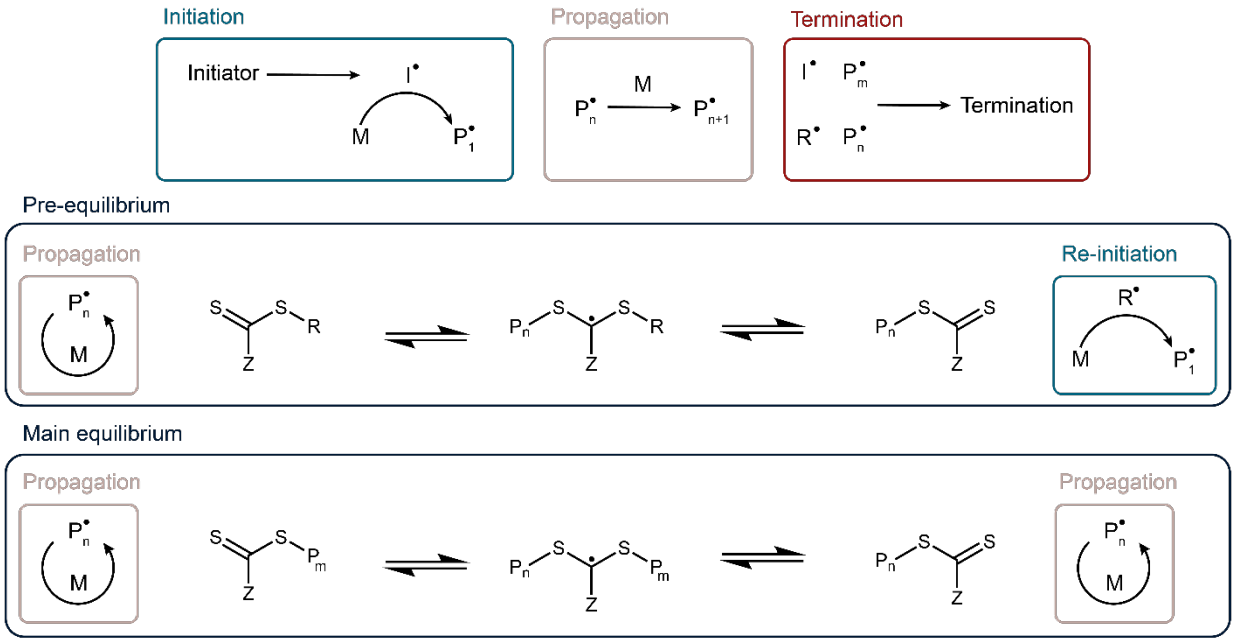


Figure 1-9: Schematic of the different steps of reversible addition-fragmentation chain-transfer (RAFT) polymerization

When the control over the primary polymer is absolute, a suitable term is Sequence-Defined polymer. In contrast, when the sequence is controlled to some degree, the term Sequence-Controlled is often used in literature.

Optimising strategy one: MIP-approach

Molecularly Imprinted Polymers (MIPs) are a synthetic equivalent to biological antibodies. The working mechanism of a MIP is best described with a 'lock and key' metaphor where the template molecule is the key and the MIP is the lock. A good lock is only effective when there is only one type of key that can open it and thus, similarly, a good MIP is only effective when it has a high affinity for

its specific template. There are various methods to synthesize MIPs: synthesis in presence of the monomer, phase inversion and soft lithography. In this section, the focus will be on the synthesis method since that is the most relevant for later applications. The MIPs are created by polymerizing monomers and cross linker in presence of the desired target molecule. After extraction of the template, highly selective cavities are left in the polymer matrix (Figure 1-10). These cavities can rebind only its specific target molecule, even when similar analogue molecules are present.[31] Commercially, most MIPs are used for separation but there are techniques to integrate MIPs into sensors. Commercially, MIPs are attractive because of their stability, low production costs, long shelf life and easy synthesis.[31]

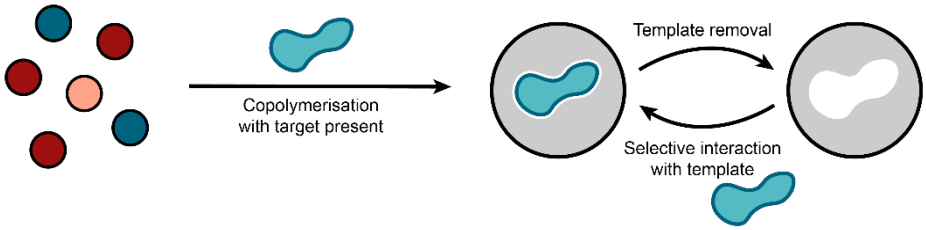


Figure 1-10: General scheme of the synthesis and working of a molecularly imprinted polymer.

The synthesis of polymer solubilizers via the 'MIP-approach' is to polymerize selected monomers in presence of the target molecule to create a 'linear MIP'. It has to be noted that the term 'MIP-approach' is chosen because the main strategy is based on the synthesis method of the MIPs. The final solubilizers won't be MIPs. Theoretically, sequences that have a high affinity with the target drug, and are statistically present in a normal copolymer, should occur more when using the MIP approach.

Optimising strategy two: the SUMI-approach

The multistep growth synthesis is a strategy to generate controlled nano-scale molecules. In here, one monomer is added at the time and this technique is used in both linear iterative synthesis and dendrimer synthesis. Single Unit Monomer Insertion (SUMI) falls under both the chain growth polymerization and the multistep growth polymerization categories. Haven *et al.* describes SUMI reactions both in batch and in flow.[32] After each monomer addition, the polymer is purified to obtain only the oligomers with the desired sequence. The term Multiple Unit Monomer Insertion (MUMI) is more appropriate when more than one monomer is built in to the sequence.

When it comes to sequence defined materials, nature has perfected its techniques over millennia of evolution. The body is constantly copying or creating specific sequences to build DNA strings or form proteins. Even though synthetic chemistry has its advantages, when it comes to sequence-defined materials it has not even come close to the efficiency of nature. Among others, these specific sequences can contain information and/or influence the structure of the final molecule. A single

variation in the sequence, often described as a point mutation, can have severe consequences such as sickle-cell anaemia, neurofibromatosis, and even cancer.[33] This illustrates the importance of the sequence, and thus for some applications, sequence-controlled molecules are not sufficient. As mentioned in **section 2**, multistep growth synthesis is a technique whereby one monomer is added at a time which creates a defined sequence. These sequence-defined techniques are harder to optimize but ultimately give more control over the final polymer. What increases the 'playability' of synthetic polymer possibilities is the fact that synthetic chemistry has a bigger variety of monomers whereas nature is limited to its specific building blocks. Aerst et al. documented a SUMI of N,N-dimethylacrylamide with RAFT polymerization in an aquatic environment. The RAFT-agent ((((2-carboxyethyl)thio)carbonothiotl)thio)-4-cyanopentanoic acid was used and it was discovered that red light provided slower but more selective excitation.[34]

Maron *et al.* demonstrated how synthetic solubilizers can be made by learning from peptide sequences.[14] By imitating the peptide properties, specific sequences were made and tested on meta-tetra(hydroxyphenyl)chlorin. Some solubilizers reached 40% higher payloads and 27-times faster initial drug release.

Optimising strategy three: data driven correlation

The problem with the SUMI-approach is that determining the optimal sequence takes a lot of time and effort. Characterising every single combination, even from a limited combination of monomers, is an impossible task to do by hand. Instead of filling out everything, machine learning could be used as a tool to predict unknown knowledge based on known data. For example, Koide et al. describes how they use machine learning to predict the cardiac dose in breast cancer radiotherapy.[35] Ghanzouri et al. developed an automatic tool to detect peripheral artery disease.[36] In another example, several machine learning techniques where compared to predict COVID-19 severity.[37]

Thus, instead of relating properties to monomers via physical experiments, a fresh look is to correlate its properties via statistical methods.

3. Machine learning

The relation between us and machines has fascinated humanity for a long time. The history of the computer goes back to the 19th century, but only gained traction around the Second World War. The first machine which could be described as a mechanical computer was invented by Charles Babbage in 1822 and was used to make tables for polynomials.[38] In the Second World War, Alan Turing's work on creating a machine to decrypt the messages of the Germans made him seen as the father of modern computing. [39-41] After the war, in 1946, the first fully programmable, electronic, general-purpose computer was unveiled at the University of Pennsylvania: The Electronic Numerical Integrator and Computer (ENIAC).[40] ENIAC was several orders of magnitude faster than its mechanical predecessors.

In 1950 Alan Turing wrote the famous article “Computing Machinery and Intelligence”, where he discussed the question “Can machines think?”. A machine or Artificial Intelligence (AI) passes the Turing Test if it is indistinguishable from a human during a conversation with a real person.[38] In the same year, Isaac Asimov, known for his three laws of robotics, bundled a series of short stories in "I, Robot".[42] The stories revolve around human-robot interactions and morality, a concept that decades later is still relevant.

Nowadays, the field of Artificial Intelligence encompasses more than just conversation: AI algorithms are responsible for the analysis of enormous datasets, predicting results given prior information on a topic, or creating images, sounds and news stories. Therefore, we give another definition to AI: "AI refers to those artificial agents capable of analysing the environment and taking actions, emulating the human reasoning process, in order to achieve a specific goal.".[41] From here, we can start to explain what encompasses Artificial Intelligence and Machine Learning (ML).

Types of Machine Learning

The goal of Machine Leaning is to let a computer algorithmically teach itself to perform a certain task. Generally, we can divide the way to train an AI in three categories: (i) supervised learning, (ii) unsupervised learning, and (iii) reinforcement learning. In supervised learning, we provide the algorithm with both the inputs and the desired outputs. In unsupervised learning only inputs are given, and the algorithm needs to work out by itself how to categorise the data. Finally, in reinforcement learning the computer must learn to solve a problem, for example learning how to play a certain video game and give itself feedback on how to improve itself.[39, 43]

On the other hand, ML algorithms can be categorised by the task they are trying to solve and are related to their learning algorithm. (a) Classification, where a model is built to automatically classify new data under the correct label, and (b) regression, where the model fits data points to predict a numeric value for new input data, are categorised under supervised learning. In (c) clustering, data

is divided in groups and are characterized by high similarity within the same group, but data from different groups should be dissimilar. As no desired grouping is given from the start (unsupervised), the algorithm should figure out grouping parameters by itself. Another example of unsupervised learning is d) dimensionality reduction, where to goal is to reduce the size of the sample dataset by selecting only certain attributes or data sampling to increase the performance of the model. This list is by no means exhaustive but is sufficient for now.

The last point of dimensionality reduction is of great importance in this thesis: it leads us to the question whether having a dataset with many attributes is better than a reduced one. On the one hand, one might say that more data is more information to make a better prediction. On the other hand, datapoints can be correlated or have no influence on the outcome of the prediction whatsoever. This data will make it harder to create a performant model since more "good" datapoints are needed to suppress the noise from the bad ones. Therefore, a good choice of features and attributes is important. However, the focus in this thesis will be on (b) regression. As such, feature choice will be a manual task. Deep Neural Networks can solve the problem of manual selection, but this goes beyond the scope of the thesis.

Machine Learning workflow

Finding patterns in data becomes more complicated with systems that exist of a complex mixture of inputs. ML automates the identification of patterns via computational calculations. ML finds its base in statistics and thus (an important factor to keep in mind when it comes to machine learning) is that at its core, it can only be as good as the provided data. Fundamentally, a ML model or algorithm is trained on existing data which then can be applied to predict an unknown entry. There are two phases: training the algorithm and evaluation. A good way to explain is using an example shown by Strieth-Kalthoff et al. (Figure 1-11).[44] In this figure, the data points (in this case monomers) are encoded with a number of specific features x_1, x_2 and x_3. Each entry has a specific target y and the ML model uses the data point, features and targets to train. After the training, the model is used to predict the target y of a new monomer with known features. Once the training phase is complete, the model needs to be evaluated to see if the value of predicted target is in the range of it's expected value and is non-biased. A big difficulty with ML is when the model learns the noise in the data as well, resulting in overfitting.

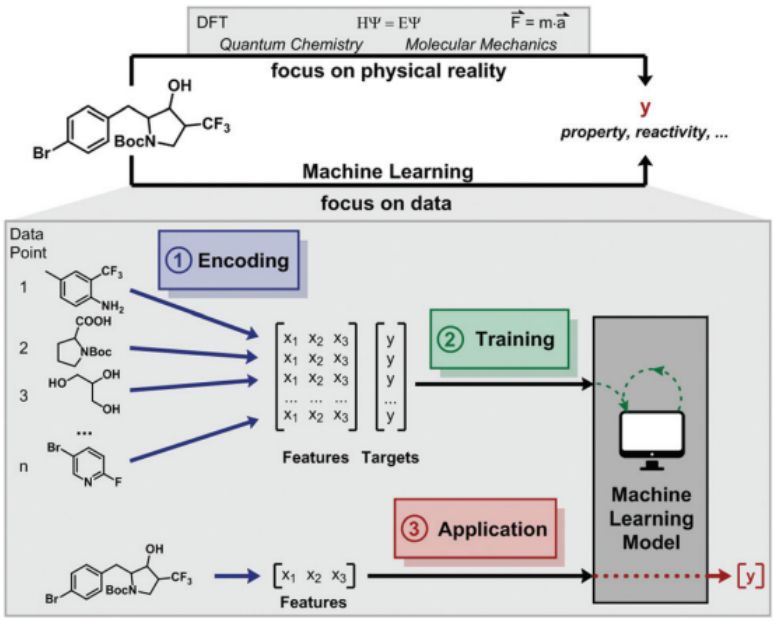


Figure 1-11: General workflow for building and applying a supervised, machine learning models for predictions. The insert figure is reproduced from Ref[44]

Regression models

There are various different algorithms and models. One of the most straightforward ones is a nested cross-validation. With a nested cross-validation, a machine learning model is chosen and the average of multiple splits is calculated before doing a performance evaluation. This machine learning model can be different things, three of which will be explained further: multivariable linear regression, Ridge regression and Least Absolute Shrinkage and Selection Operator (LASSO) regression.

Multivariate linear regression (henceforth linear regression) is a combination of several linear regressions on independent variables for one dependant variable (Equation 6) [45, 46]:

$$Y = XB + U \qquad \text{Equation 6}$$

where Y is a matrix of the measurements, and X are descriptors of Y (labels and features, respectively, in machine learning terms). B contains the weights assigned to each feature in X, and U is any remaining noise and error. The algorithm determines the optimal values of B, by for example minimising the sum of the squared residuals.

Ridge regression is very similar to a linear regression with the difference of an added penalty term also referred to as *L2*. The penalty is equal to the squared value of the coefficients. *L2* regularisation searches for the minimum of the sum of the squared coefficients by tuning a tuning parameter, λ, in the following cost function (Equation 7). In this equation, y_i is the i-th datapoint, w_j is the weight, x is the feature value, and λ is the penalty term.[47, 48]

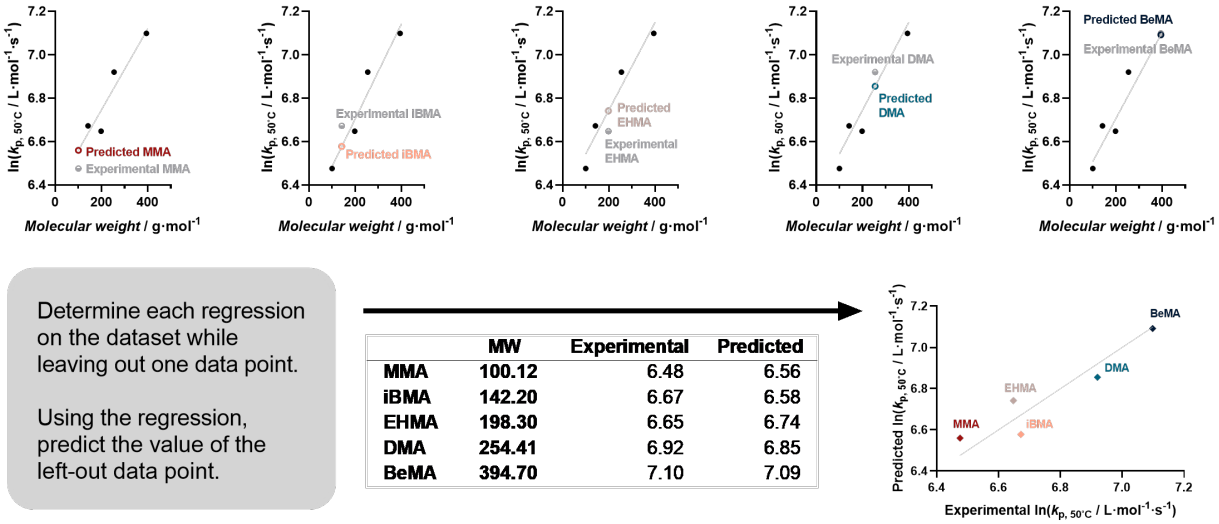
$$\min\left(\sum_{i=1}^{M}(y_i-\hat{y}_i)^2+\lambda\sum_{j=0}^{p}{w_j}^2\right)=\min\left(\sum_{i=1}^{M}(y_i-\sum_{j=0}^{p}w_j*x_{ij})^2+\lambda\sum_{j=0}^{p}{w_j}^2\right)$$ Equation 7

Lasso regression is like ridge regression; however, the cost function instead minimises the absolute sum of coefficients, known as *L1* regularisation (Equation 8). As a result, lasso regression not only assists to reduce overfitting, but it can be used in feature selection.[49]

$$\min\left(\sum_{i=1}^{M}(y_i-\hat{y}_i)^2+\lambda\sum_{j=0}^{p}{w_j}^2\right)=\min\left(\sum_{i=1}^{M}(y_i-\sum_{j=0}^{p}w_j*x_{ij})^2+\lambda\sum_{j=0}^{p}|w_j|\right)$$ Equation 8

Since λ is scale variant, and thus requires continuous data do be standardised for both Ridge and LASSO regression. Additionally, one must bear in mind that when λ increases, the bias increases, but the variance decreases.[49]

It is useful to know that in case of a limited amount of data entries, the Leave-One-Out-Cross-Validation (LOOCV) is something to consider (Figure 1-12).[50, 51] In this illustrative scheme, there are only five data points. Thus, five regressions were executed whereby every time a different data point was not included in the regression. These regressions are used to predict the 'missing' data point which subsequently gets compared to its correct value to evaluate the algorithm. The closer the predicted value is to its experimental value, the closer it will be to the x = y line.



	MW	Experimental	Predicted
MMA	**100.12**	6.48	6.56
iBMA	**142.20**	6.67	6.58
EHMA	**198.30**	6.65	6.74
DMA	**254.41**	6.92	6.85
BeMA	**394.70**	7.10	7.09

Figure 1-12: Example of the Leave One Out Cross Validation Method (LOOCV) where the LOOCV method is applied to five data points using a linear regression.

Something that is a bit more complex than a simple regression is a decision tree or random forest. This technique works by determining a list of Boolean decisions where each feature of significance influences the final prediction of the tree model (Figure 1-13. They operate simpler than a neural network, and there are methods to visualize the decision-making process that a decision tree will undergo to reach its final prediction.[46]

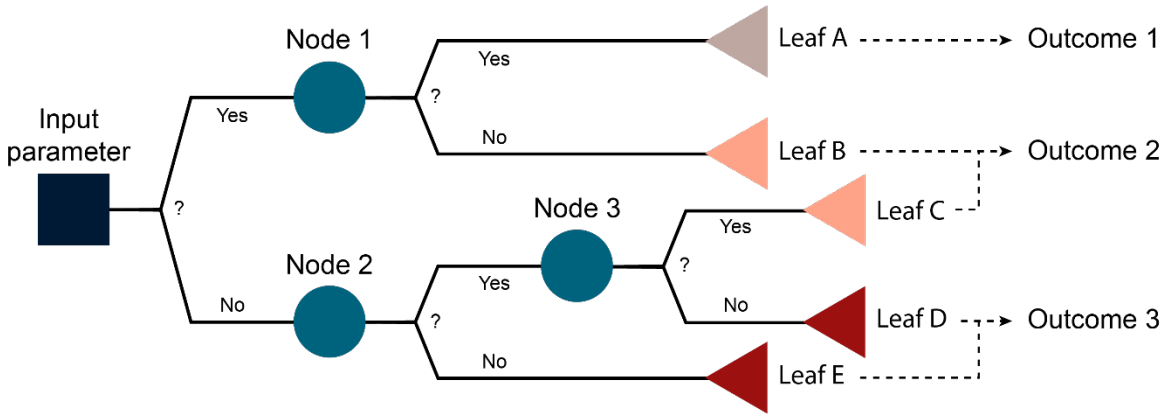


Figure 1-13: Visualisation of a decision tree where from the input (dark blue square) “branches” (black lines) divide itself over multiple nodes (blue circle) until its final “leaf”. Several leaves can result into the same outcome.

A neural network is more complex than a decision tree since the inner workings and resulting models are not necessarily understood (Figure 1-14). Unfortunately, they tend to overfit the model, especially with a limited dataset.[27, 52]

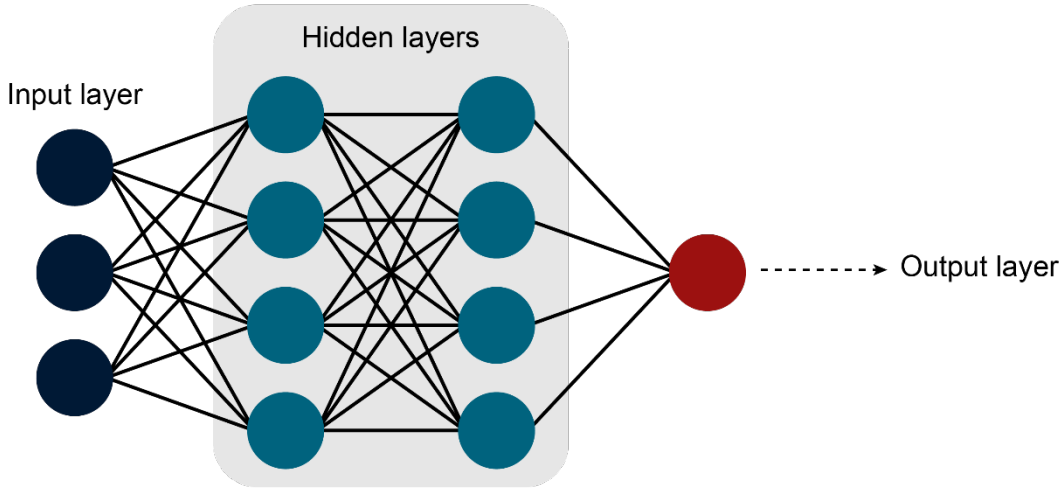


Figure 1-14: Visualisation of a Neural network where a more complex “hidden layer” connects the input layer and the output layer.

Recent developments

Artificial intelligence gains more and more popularity over the years. As illustrated in Baum et al., the percentage of publications that have AI-related topics skyrockets between 2015 and 2020. Especially in analytical chemistry, the use of AI becomes more and more popular.

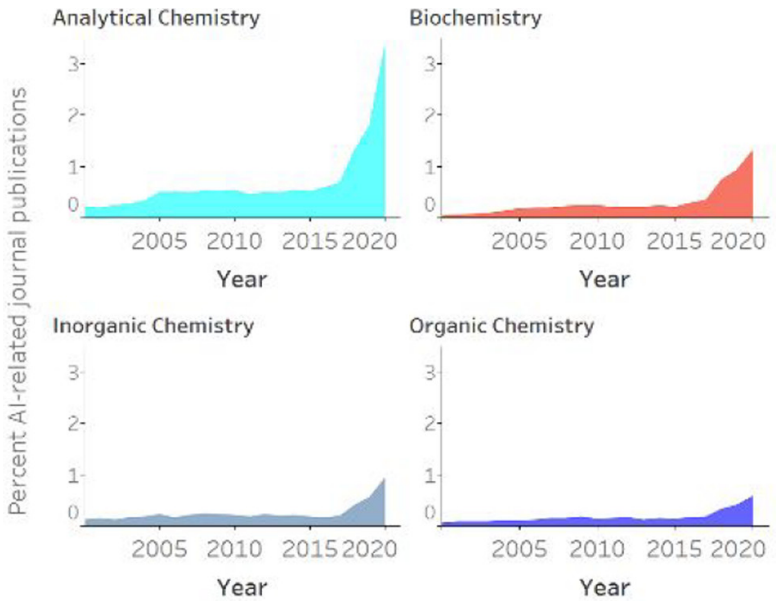


Figure 1-15: Percentage of Artificial Intelligence (AI) related journal publications for different types of chemistry. The insert figure is reproduced from Ref[53]

Several interesting techniques and discoveries are made with the help of AI. For example, Raccuglia et al. made a machine learning model, based on data from unsuccessful reactions, that predicts the success rate of new reaction outcomes.[54] Because journals often omit the experiments that were not successful, a big chunk of data never is revealed to the "outside world". Using failed results to help predict better conditions for future experiments is a clever method to improve overall success rate. Other works, such as that from Ramakrishnan et al., focus more on strategies that reduce the computational cost of quantum chemistry.[55] Coley et al. developed a robotic platform for flow synthesis of organic compounds.[56] By combining both intelligence-driven synthesis planning and robotically controlled experiments, they successfully synthesized fifteen drug (or drug like) substances.

4. General aims

The research will aim to combine two different branches of science: polymer chemistry and computer science. The first chapter starts with a "classic" chemistry project where tailored polymers for a specific drug delivery application are investigated. Chapter 2 and chapter 3 steer away from the more conventional lab experiments and explore the possibilities of machine learning as a tool to help find more efficient chemistry. All these chapters are only a couple of pieces in a much larger puzzle, which is to use machine learning as a tool to have more streamlined science. However, an added bonus of delving into machine learning is that it can be done remotely, useful in a global pandemic.

Chapter 2: Enhancing the Efficiency of Statistical Copolymers towards Drug Solubility Applications

1. Introduction

In the biological world, sequence-defined macromolecules such as proteins or DNA are mayor building blocks for life. Nature has evolved to create complex pathways to synthesise a specific molecule for a desired application. As such, nature has often been an inspiration in applied materials science. Although synthetic macromolecules are not as sophisticated as their biological counterparts yet, they do have their advantages. First, there is a larger variety in synthetic building blocks as opposed to the ones that occur in nature, leading to a larger variety in chemical compounds. Secondly, stability is a mayor issue when it comes to for example proteins. The right synthetic counterpart could improve shelf-life and reduce costs for storage and transportation.[14]

As stated before (Chapter 1), hydrophobic drugs cause an additional challenge in drug delivery inside the body.[3, 4] Using copolymers to enhance the solubility is not a new idea, as there are already existing products on the markets. Kanikkannan et al. lists different examples ranging from solid dispersion based products to lipid based delivery systems that are produced by companies like Janssen, Merck and Pfizer.[57] Polyvinyl alcohol (PVA)-based and gelatin-based polymers were recently (2022) investigated by Choi et al. to see the effect on Rivaroxaban.[58] The commercially available dimethylaminoethyl methacrylate-copolymer Eudragit EPO (EPO) was investigated by various people such as Saal et al. (2018)[59] and Fine-Shamir et al. (2019).[60] To the authors knowledge, there does not seem to be research done about the specific sequence of these copolymers except in the work of Maron et al.[14] They mirrored the functionality of a peptide into the side-chain function of a synthetic copolymer. Taking inspiration from this, it was theorised that there must be sequences, tailored to a target drug molecule, that are more effective solubilizing said drug.

The search for a perfect tailored polymer for the right application is unfortunately a lengthy and costly process. In theory, there is an ideal sequence of monomers that interacts with the desired target molecule to bring it into solution. However, when a random copolymer using the same starting molecules is synthesized, that ideal sequence will occur naturally throughout the polymer. When the polymerisation occurs in presence of the target molecule, the idea is that the monomers themselves will orient themselves in the desired sequence since that position is the most stable.[31] The idea of this imprinting process is derived from Molecularly Imprinted Polymers (MIPs) (Figure 2-1).

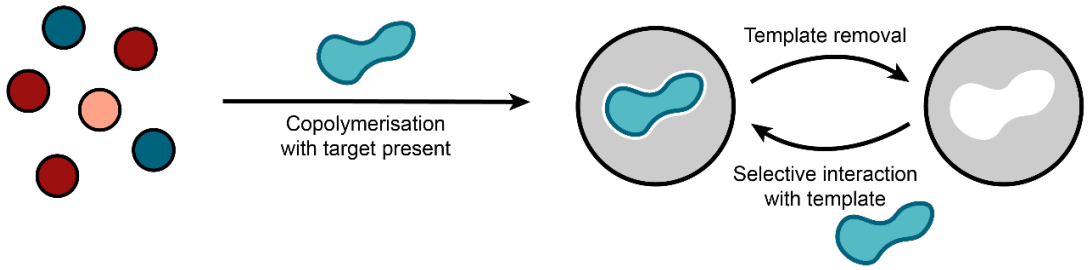


Figure 2-1: Theory of Molecularly Imprinted Polymers

After removing the original template molecule, the target can selectively rebind in the precise cavity of the MIP.[31] However, in this work there will be no cavity since there won't be any use of cross-linker. Rather, a "linear MIP" is a better way to describe our final copolymers. Originally, the idea was to compare the efficiency of the "linear MIPs" (the statistical copolymers) with the efficiency of a tailored sequence (defined copolymers) (Figure 2-2). Unfortunately, this part of the project was swapped for something more suitable in a COVID-19 environment and is thus not discussed.

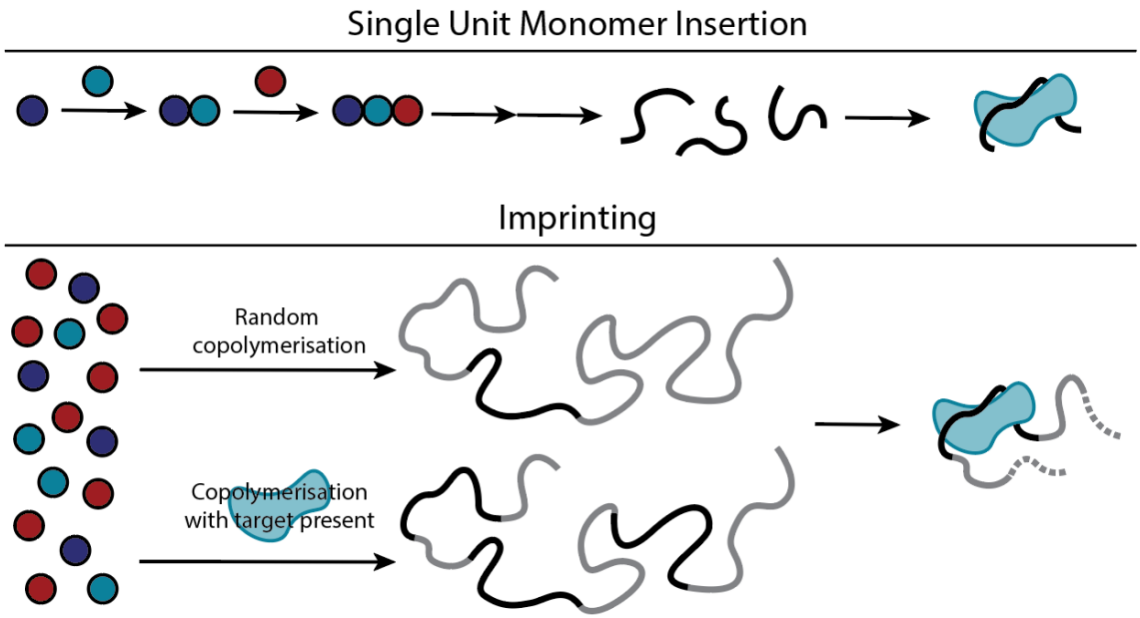


Figure 2-2: Comparison between the polymer interacting with their target for (1) the Single Monomer Insertion synthesis technique and (2) the Imprinting technique.

In this chapter, methyl acrylate copolymers are investigated as potential solubilizers as proof of principle for the nonsteroidal anti-inflammatory drug Piroxicam (PCX). Saal et al. demonstrated that the commercially available Eudragit® EPO can be used as a solubilizer for a variety of different hydrophobic drugs (Figure 2-3b).[59] One of such hydrophobic drugs is the nonsteroidal anti-inflammatory Piroxicam (Figure 2-3c).

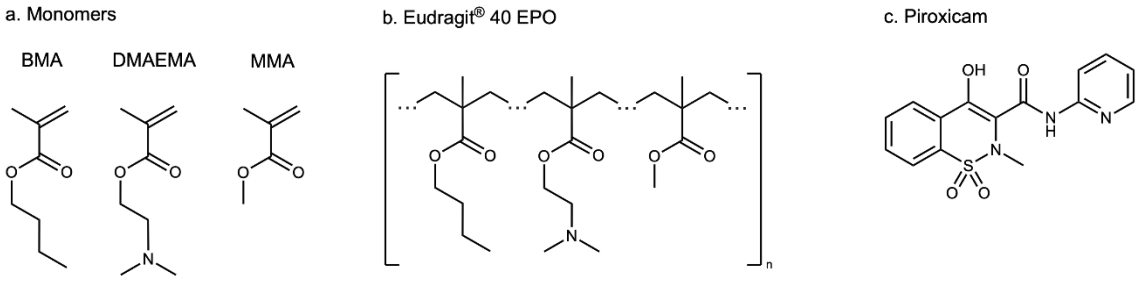


Figure 2-3: Structure of butyl methacrylate (BMA), 2-(dimethylamino)ethyl methacrylate (DMAEMA) and methyl methacrylate (MMA), Eudragit® 40 EPO and Piroxicam

The copolymers are based on the Eudragit® EPO using butyl methacrylate (BMA), 2-(dimethylamino)ethyl methacrylate (DMAEMA) and methyl methacrylate (MMA) as monomers (Figure 2-3a). However, rather than trying to finds a good solubilizer sequence by trial and error, it was opted to find the correct sequence via imprinting (Figure 2-2). Eudragit® EPO was chosen since it both is commercially available and straightforward to synthesise.

2. Proof of complete random copolymerisation

As explained in the introduction, free radical polymerisation exists of three steps: initiation, propagation and termination (Figure 1-4). All the polymerisations were executed with azobisisobutyronitrile (AIBN) as initiator and a methacrylate as monomer (Figure 2-4).

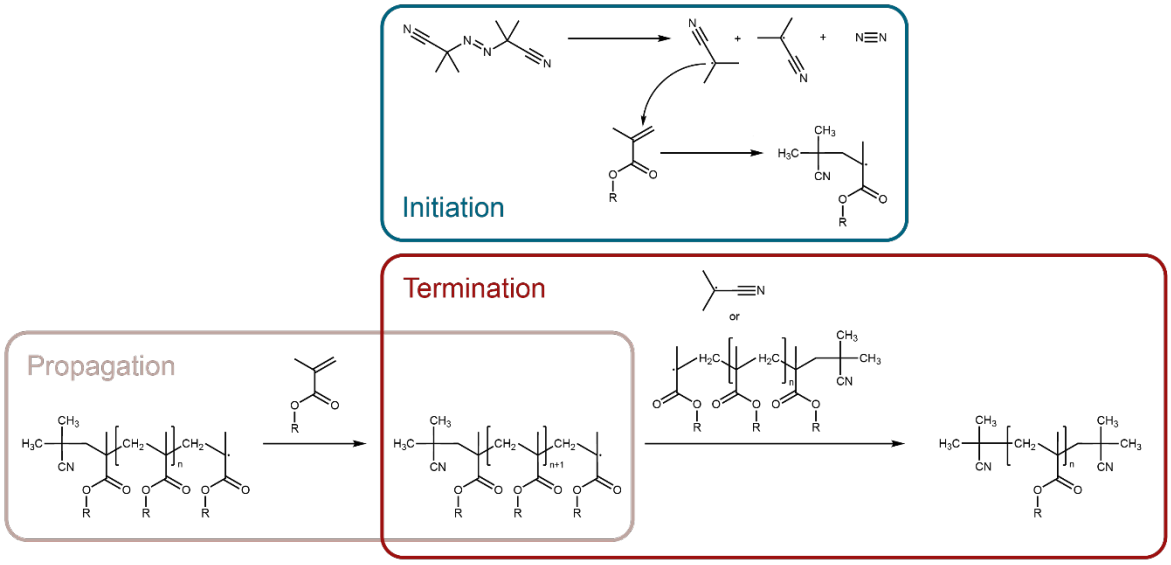


Figure 2-4: Initiation (blue), propagation (grey) and termination (red) of a free radical polymerisation using azobisisobutyronitrile (AIBN) as initiator and a methacrylate as monomer.

In order to ensure the copolymerisation of the three monomers, BMA, DMAEMA and MMA, was completely random, a series of tests were performed. First, six different polymerizations were executed in bulk, using AIBN (0.0390 mmol) and a mixture of the three different monomers (7.8 mmol). The ratio BMA:DMAEMA:MMA was subsequently 1:1:0, 1:0:1, 0:1:1, 1:1:1, 1:2:1 and 1:3:1. The mixture is polymerized at 70 ˚C for 24 hours after purging with argon.

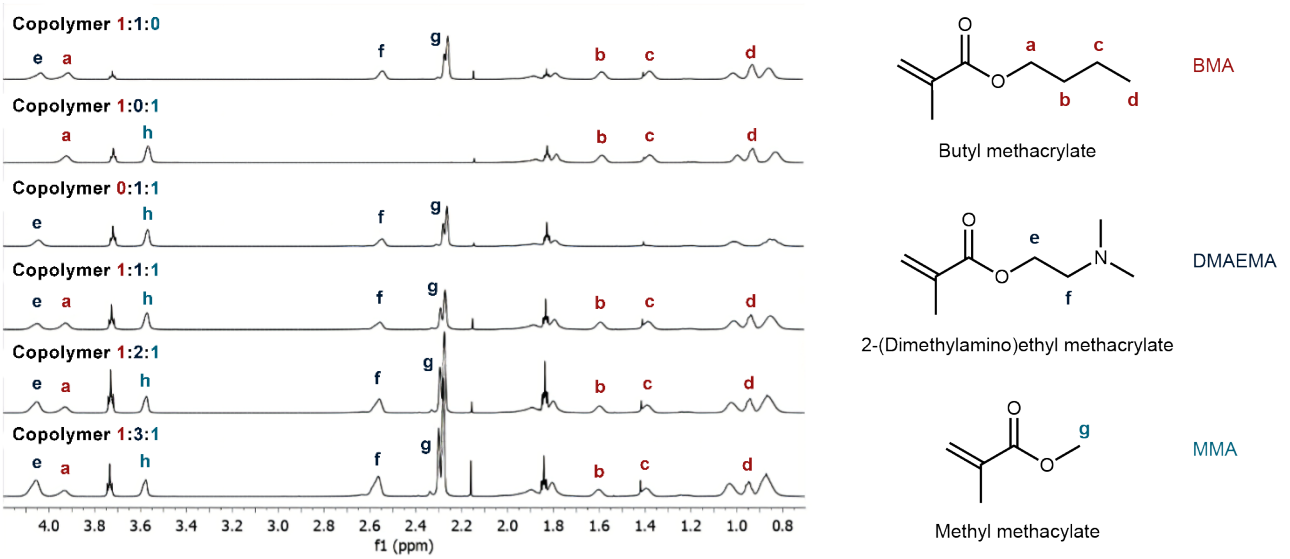


Figure 2-5: ^{1}H-NMR (300 MHz, $CDCl_3$) of six different copolymers with a different ratio of butyl methacrylate (BMA, red), methyl methacrylate (MMA, green) and 2-(dimethyl)aminoethyl methacrylate (DMAEMA, blue). Subsequently, the ratio BMA:DMAEMA:MMA was 1:1:0, 1:0:1, 0:1:1, 1:1:1, 1:2:1 and 1:3:1. The mixture is polymerized at 70 ˚C for 24 hours after with argon. After reaction, the copolymer was isolated by precipitation in petroleum ether and dried.

After the reaction, the copolymer was isolated by precipitation in petroleum ether and dried. 1H-NMR (300 MHz, CDCl3) results indicate that the difference of the BMA:DMAEMA:MMA ratio of the added monomers and the ratio of the BMA:DMAEMA:MMA build into the polymer was less than 5% (Figure 2-5, Table 1).

Table 1: Free radical polymerisations using of butyl methacrylate (BMA), 2-(dimethylamino)ethyl methacrylate (DMAEMA) and methyl methacrylate (MMA) for 24 hours at 70 ˚C in batch (no solvent). The ratio between the monomers was determined by weighing the amount of each monomer added to the stock solution. The ratio of the monomers built in in the polymer is determined with H^1 NMR after the polymerization.

	Targeted ratio in polymer			**Feed ratio monomers**			**Build-in polymer**		
	BMA	DMAEMA	MMA	BMA	DMAEMA	MMA	BMA	DMAEMA	MMA
1:1:0	1	1	0	1	1.00	0	1	1.00	0
1:0:1	1	0	1	1	0	1.00	1	0	0.99
0:1:1	0	1	1	0	1	1.03	0	1	1.03
1:1:1	1	1	1	1	0.98	1.03	1	1.02	1.03
1:2:1	1	2	1	1	1.95	1.05	1	1.92	1.01
1:3:1	1	3	1	1	2.86	1.05	1	2.79	1.01

The build-in polymer ratio was determined by normalising the 1H-NMR peak at 4 ppm, which is assigned to polymer BMA to two (Figure 2-6, BMA'). The ratios were subsequently calculated by dividing the polymer BMA peak (Figure 2-6, BMA') by two, the polymer DMAEMA peak (Figure 2-6, DMAEMA') by four and the polymer MMA peak (Figure 2-6, MMA') by three. The 5% difference with the feed ratio of monomers is negligible since the 1H-NMR has an accuracy of 5%.

Second, a 1:2:1 BMA:DMAEMA:MMA ratio was polymerized in different solvents. Six different polymerizations were executed in six different solvents, using AIBN (0.0213 mmol) and a mixture of three different monomers (4.26 mmol). The monomer mixture and the AIBN were dissolved in subsequently chloroform, dichloromethane, dimethylformide, dioxane, tetrahydrofuran and toluene. Similarly, the mixture was polymerized at 70 ˚C for 24 hours after with argon.

Table 2: Free radical polymerisations using butyl methacrylate (BMA), 2-(dimethylamino)ethyl methacrylate (DMAEMA) and methyl methacrylate (MMA) for 24 hours at 70 ˚C in different solvents. The abbreviations are as followed: $CHCl_3$ = chloroform, DMF = dimethylformamide, $C_8H_{16}O_4$ = dioxane, THF = tetrahydrofuran, MePh = toluene. The ratio between the monomers was determined by weighing the amount of each monomer added to the stock solution. The ratio of the monomers built in in the polymer is determined with H^1 NMR after the polymerization.

	Targeted ratio in polymer			**Feed ratio monomers**			**Build-in polymer**		
	BMA	DMAEMA	MMA	BMA	DMAEMA	MMA	BMA	DMAEMA	MMA
$CHCl_3$	1	2	1	1	1.89	0.97	1	1.83	0.94
DMF	1	2	1	1	1.89	0.97	1	1.84	0.94
$C_8H_{16}O_4$	1	2	1	1	1.89	0.97	1	1.84	0.93
THF	1	2	1	1	1.89	0.97	1	1.83	0.92
MePh	1	2	1	1	1.89	0.97	1	1.86	0.91

Again, the difference between the added monomer ratio and the ratio in the polymer is negligible (Table 2). In conclusion, that the tested solvents don't have a noticeable influence on the BMA:DMAEMA:MMA ratio in the final polymer.

To exclude the possibility that the copolymer is not random, the polymerization was stopped at low conversion. The same polymer mixture, 0.462 mmol monomer with a ratio 1:2:1 ratio BMA:DMAEMA:MMA and 0.038 mmol AIBN in toluene, was polymerized at 70 degrees for 1 hour and 5 minutes.

Table 3: Free radical polymerisations using butyl methacrylate (BMA), 2-(dimethylamino)ethyl methacrylate (DMAEMA) and methyl methacrylate (MMA) for 1 hours and 5 min at 70 ˚C in different solvents. The ratio between the monomers was determined by H^1 NMR before polymerization. The ratio of the monomers built in in the polymer is determined with H^1 NMR after the polymerization.

	Targeted ratio in polymer			**Feed ratio monomers**			**Build-in polymer**		
	BMA	DMAEMA	MMA	BMA	DMAEMA	MMA	BMA	DMAEMA	MMA
1 hour	1	2	1	1	1.92	0.96	1	1.94	0.94
5 min	1	2	1	1	1.92	0.93	1	1.93	0.94

1H-NMR illustrates that even if the conversion is low, the ratio of BMA:DMAEMA:MMA stays the same no matter how small the polymers are (Figure 2-5, Table 3).

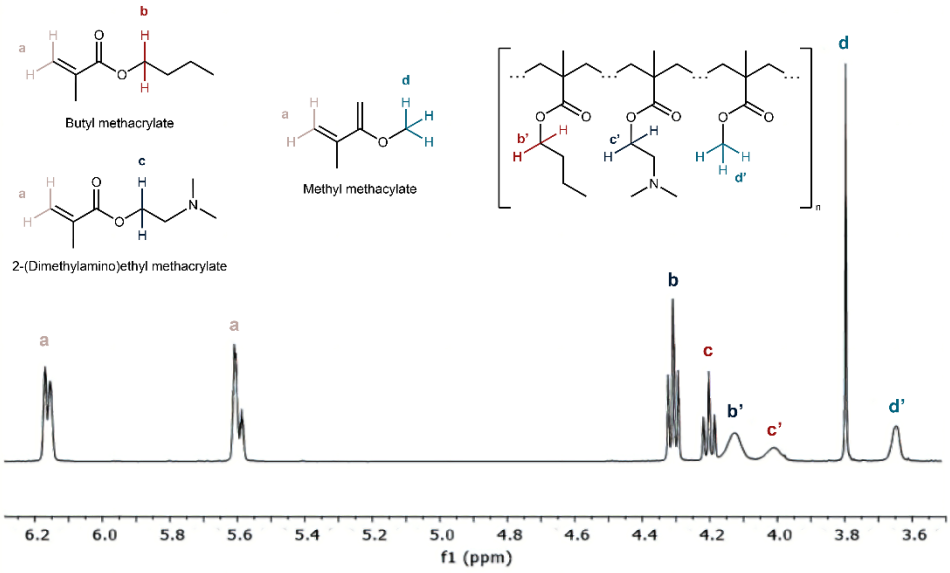


Figure 2-6: Part of the ^{1}H-NMR (300 MHz, $CDCl_3$) spectrum of a crude mixture of a polymerization using butyl methacrylate (BMA, red), methyl methacrylate (MMA, green) and 2-(dimethyl)aminoethyl methacrylate (DMAEMA, blue). A 1:2:1 ratio of BMA:DMAEMA:MMA was used and the mixture was polymerized at 70 ˚C for 1 hour. The vinyl peaks of the monomers are coloured yellow and the hydrogen atoms of the polymer and the monomer are labelled with and without an apostrophe respectively.

Unfortunately, the presence of DMAEMA in the polymer seems to have an interaction with the analytical columns of the SEC-GPC. For this reason, no data of the SEC-GPC is provided since the outcome is most likely biased.

3. Dissolving the copolymer and determining coefficient

Saal et al. mentioned that they dissolved the Eudragit® EPO directly in water.[61] However, the 1:2:1 polymer does not seem to dissolve in Demi water. Saal et al. prepared their polymer solutions by dissolving different amounts of Eudragit® EPO in deionized water and adjusting the pH to 6.0. Despite this, the 1:2:1 BMA:DMAEMA:MMA polymers had a hard time dissolving in water. However, Palena et al. investigated self-organized drug-interpolyelectrolyte nanocomplexes where they used Eudragit® EPO as well. In here, HCl was added to obtain an initial partial neutralization of the amine groups which stimulates hydration and relaxation of the polymer. After the solid dispersions were formed, the solution was neutralized with NaOH. In order to generate a more uniform pH of 7.4 over all the solutions, a polyphosphate buffer was used instead of water to dissolve all the polymer into a hydrophilic substance. The presence of Piroxicam was determined via UV-Vis spectroscopy since the copolymer does not show any absorbance (Figure 2-7).

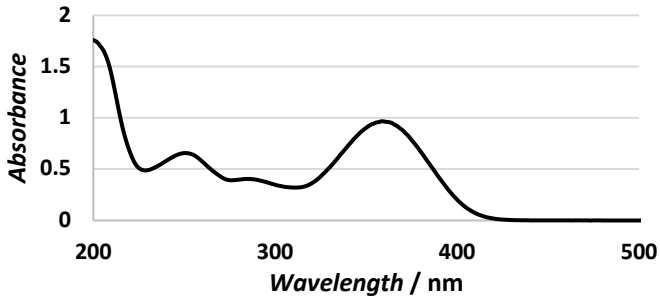


Figure 2-7: Example of UV-Vis spectrum of Piroxicam dissolved in a phosphate-buffered saline (PBS) buffer using a 1:2:1 ratio of BMA:DMAEMA:MMA copolymer

The peak of the wavelength, 360 nm, is used to determine the absorbance of future UV-Vis spectra. A dissolution series of Piroxicam, where the drug was dissolved into a 2wt% 1:2:1 BMA:DMAEMA:MMA copolymer in a phosphate buffered saline (PBS) buffer, was determined (Figure 2-8).

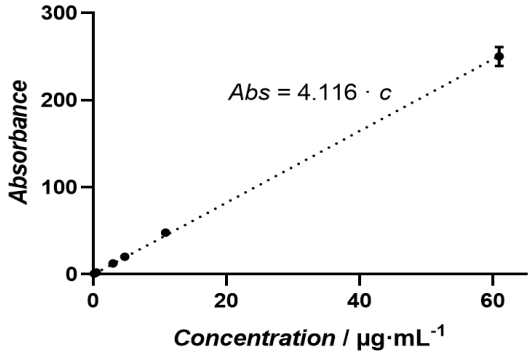


Figure 2-8: Determination of Molar Absorption Coefficient of Piroxicam in a 2 wt% 1:2:1 BMA:DMAEMA:MMA copolymer in polyphosphate buffer

For future calculations, the absorbance can be correlated to 4.115 times the concentration in micrograms per millilitre.

4. Determining the optimal solubility enhancement

As mentioned before, the 1:2:1 BMA:DMAEMA:MMA copolymer is tricky to dissolve on its own in water or buffer without acidifying the solution first. To determine the maximum solubility of Piroxicam in a given solution, the same procedure was executed every time (Figure 2-9).

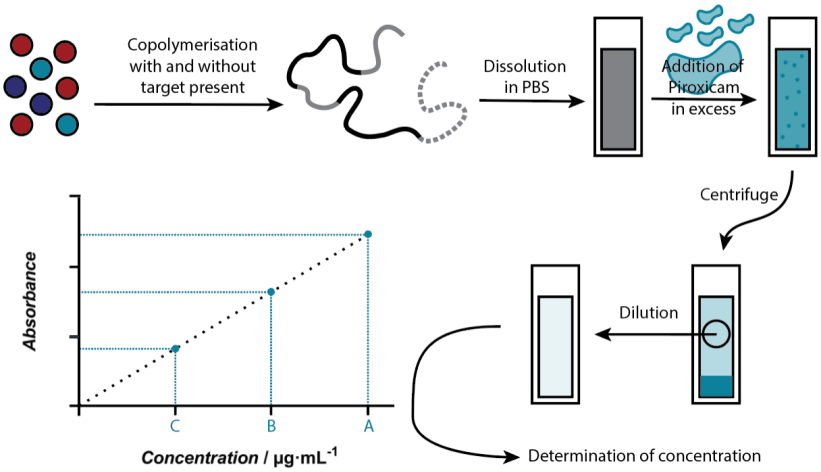


Figure 2-9: Procedure to determine the maximal solubility of a given polymer

After a copolymerisation and dissolution of the desired polymer, an excess of Piroxicam was added to 1 mL of co-polymer solution. The mixture was shaken well before centrifuging for an hour. The excess Piroxicam sinks to the bottom as a pellet and some of the supernatant was carefully extracted. Before measuring this on the UV-Vis, the mixture was diluted with PBS buffer. Using the coefficient from Figure 2-8, the original maximum concentration was determined.

To determine which polymer concentration was optimal, the maximum solubility was determined for three concentrations of the same 1:2:1 BMA:DMAEMA:MMA copolymer: 0.5 wt%, 1 wt% and 2 wt%. A higher concentration was attempted as well but proved to be too difficult to properly dissolve the polymer into PBS (Figure 2-10).

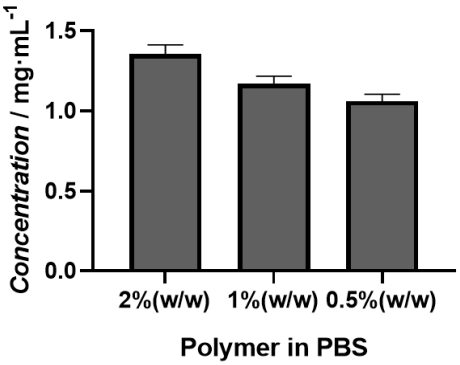


Figure 2-10: Maximum amount of Piroxicam concentration in different weight percentage solutions of 1:2:1 BMA:DMAEMA:MMA copolymer

The maximum solubility of Piroxicam varied between 1 mg·mL^{-1} and 1.5 mg·mL^{-1} with an increase in Piroxicam dissolution the higher the polymer concentration. This indicates that the amount of polymer present in the solution does have a positive effect on the overall Piroxicam solubility. In Saal et al., they did succeed in getting higher polymer concentrations but they indicated that there was a plateau phase of Piroxicam solubility after 2 wt%. For the following experiments, a concentration of 2 wt% was used. In Schmied et al. they investigated a modified version of EPO as well, only a controlled molecular weight. It would have been interesting to vary the molecular weight of the polymer to see its effect on the solubility enhancement but this was unfortunately not pursued due to COVID-19 restrictions.

The 1:2:1 BMA:DMAEMA:MMA copolymer was synthesised in presence of Piroxicam as well. Per monomer unit, a ratio of 1.75 units of Piroxicam was mixed into the polymerisation mixture. 0.462 mmol monomer with a ratio 1:2:1 ratio BMA:DMAEMA:MMA and 0.038 mmol AIBN in tetrahydrofuran, was polymerized at 70 degrees for 1 hour. After polymerisation, an attempt was made to purify the polymers and remove all of the Piroxicam (Figure 2-11).

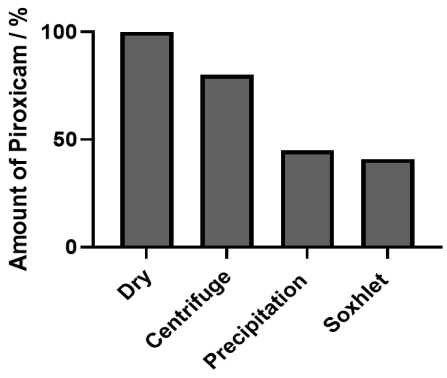


Figure 2-11: Percentage of Piroxicam left for several techniques.

The reference for extraction techniques is letting all the solvent evaporate using a rotary evaporator. The amount of Piroxicam left by this technique is set to 100% since it assumes that no Piroxicam is

lost. In the first test, the polymer mixture was centrifuged in order to let all the Piroxicam sink to the bottom. This results in a reduction of around 20%, indicating that the polymer interaction was strong enough to keep it into solution. Second, and third, precipitation and Soxhlet extraction using petroleum ether were executed. The Soxhlet extraction was executed overnight. Both techniques were able to remove around half the Piroxicam but not completely purify it.

Different 1:2:1 BMA:DMAEMA:MMA copolymers were synthesised in presence of Piroxicam. Per monomer unit, four different units of Piroxicam, 0.00, 0.75, 1.25 and 1.75, were mixed into the polymerisation mixture. The mixtures were polymerized at 70 degrees C for 1 hour. Despite the not complete purification technique, precipitation in petroleum ether was chosen because of its time efficiency. The maximum solubility of Piroxicam for each copolymer was determined and plotted compared to the literature (Figure 2-12).

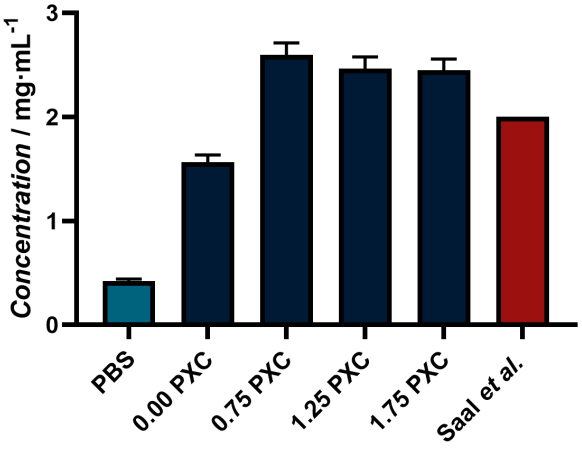


Figure 2-12: Amount of maximum Piroxicam solubility for different copolymers (blue) compared to literature (red)

There is a significant increase whether or not the original polymer was synthesised in presence of Piroxicam. Compared to the literature, “empty” polymers were less effective whereas “imprinted” polymers reached a solubility of 2.5 mg·mL^{-1}. This indicates that the imprinting method improves the effectiveness of the copolymers. However, there is no increase of effectiveness if the original concentration of Piroxicam is increased. Interesting to note is that in multiple studies of Saal et al. there are a multitude of drugs investigated besides Piroxicam focussing on anionic drugs (bezafibrate, furosemide, indomethacin, etc.) and acidic drugs (pimozide, tamoxifen, carvedilol, etc.).[59, 61] This work focusses on Piroxicam as a proof of concept but the “polymerisation in presence of the drug” technique can be applied to a multitude of other drugs as well.

5. Results and discussion

Within the experiments of this chapter, an optimal 1:2:1 BMA:DMAEMA:MMA copolymer was tested as a solubilizer for the hydrophobic drug Piroxicam in PBS buffer. The optimal copolymer was synthesised in presence of its target drug in tetrahydrofuran at 70 degrees Celsius for 1 hour. The copolymer was isolated from the solvent and successfully redissolved in an hydrophilic solvent, PBS buffer. The polymerisation was, without presence of the target molecule, proved to be random via ^{1}H-NMR. It is assumed that the presence of Piroxicam influences the monomer sequence by having the monomers orientate naturally around the target molecule before polymerisation by effects such as hydrogen bonding. During polymerisation, the influenced sequence is locked in place leading to an overall more efficient copolymer. The maximum concentration of polymer to PBS buffer was 2%(w/w). Precipitation and Soxhlet extraction were the most effective methods to remove around 50% of the original the target molecule concentration. Ideally, the target molecule needs to be removed completely and more investigation is necessary to fully prove this methods effectiveness. Nevertheless, the "imprinted" polymers performed better than the polymers synthesised without Piroxicam, reaching a maximum solubility of 2.5 mg·mL^{-1}.

Other strategies to improve the solubility of Piroxicam exist. In 2005, Karataş[62] et al. improved the solubility of PXC using the two surfactants Gelucire 44/14 and Labrasol. Interesting to note is that they report a concentration of 0.0198 mg/ml in water at a temperature of 37 °C. This value matches close to our experimental values when pure water was used. However, since all the other experiments were executed in PBS buffer, it was opted to take the concentration of Piroxicam in PBS as "base" value which is larger than pure water. In Karataş' work, they report a 50-fold increase of drug solubility when using a 15% Labrasol solution which quantitively equals to 1.011 mg/mL.[62] Even without a Piroxicam imprinted polymer, this is still less than the 1.5 mg/mL we report. A more recent study of Al-Hamidi et al. in 2015 makes use of amino sugars as hydrophilic carriers.[63] Their best results were a value of 30±1.6 mg /100 mL for 10%w/v glucosamine HCl and 34.3 ±3.8 mg /100 mL for 15%w/v gluconolactone. Both values are relatively small increases compared to our findings. However, they showed no significant changes for Piroxicam using multiple solid-state studies including Fourier-Transform Infrared spectroscopy (FT-IR), Differential Scanning Calorimetry (DSC) and X-Ray Powder Diffraction (XRPD). In 2016, Patnaik et al. reports the use of nanosuspensions with Soluplus® as a Piroxicam solubiliser.[64] They reported a linear increase of drug solubility when the polymer concentration was increased with a maximum concentration just under 6.0 µg/mL or 0.006 mg/mL. In a more recent 2020 study, Ammanage[65] et al. reported the co-crystallisation and the formulation of buccan films. Their most successful finding was around four times smaller than our highest value with a reported 60.73 ±1.95 mg /100 mL for Piroxicam-sucralose film. Ammanage et al. did FTIR, DCS and X-ray diffraction (XRD) studies as well and on top of that researched the in

vitro drug release with histopathological studies.[65] Their strategy is especially interesting since they opted to create mucoadhesive films, a very different drug delivery route then most.

Naturally, this work focusses on polymers to make polymer-drug conjugates as potential solutions. However, conjugate systems do not necessarily need to exist out of polymers. A recent study of Cho et al. describes a peptide-drug conjugate whereby the use of a peptide that can cross the blood brain barrier is used to improve the drug (camptothecin) efficacy in brain cancer.[66] Thus, not only did the peptide increase the solubility of the drug in water, it also helped to reach difficult places. However, polymers can also mimic peptides as show in a study by Maron et al.[14] In this study, a well-known peptide sequence (used to solubilise the photosensitiser meta-tetra(hydroxyphenyl)-chlorin) was translated to a polymer sequence based on oligo(N-substituted acrylamide)s and oligo(2-substituted-α-hydroxy acid)s. The polymer variant of the sequence exceeded the payload and initial drug release properties, indicating a great potential in the use of polymers that mimic specific peptide sequences.

Precision polymers are an exciting potential player in biomedical applications but often are slow to produce. In order to streamline the connection between synthesis and application, and to modernise chemistry in general, this work offers a rethinking of that strategy. The initial results prove that there is a future in making these tailored polymers towards their desired target molecule. However, this work only provides the first stepping stones towards an efficient and streamlined cascade of polymer synthesis. Reflecting on the performed experiments, more efficient techniques would be used for future experiments. One of the bigger issues was the scalability and the purification of the copolymers. Dissolving the copolymer in the PBS buffer was not straightforward in and of itself and there are some options that could be considered to make this more efficient. First, the length of the copolymers could be reduced to a smaller size via a more controlled polymerisation, such as described by Schmied et al.[67] Second, the use of a specific RAFT agent itself could potentially make the polymer more hydrophilic as well. A controlled polymerisation can be done with a more hydrophilic RAFT agent or (if the RAFT agent has too much of an influence on the imprinting) the RAFT agent could be modified in a later stage to make it more hydrophilic. Nevertheless, both these strategies have various new parameters that could influence the effectiveness of the overall enhancement of the solubilizing efficiency of the copolymer. Hence, an automatised system that, potentially via machine learning, can decide quickly which parameters to vary would be the ideal future.

Thus, when the making of these copolymers becomes more streamlined, further collaborative biomedical studies (such as drug release and toxicity) can be performed to complete the story from design to working product. Future experiments, such as the effect these polymers have on the blood brain barrier and their use in delivering drugs in difficult to reach places, could open up a new future of biomedical applications.

6. Conclusion

Despite the limits of synthetic macromolecules in comparison to their more sophisticated counterparts, polymers are an interesting strategy for drug solubility. Specific, tailor made sequences are desirable, with the disadvantage that finding the right sequence often takes time and money.

In this work, a 1:2:1 BMA:DMAEMA:MMA copolymer was tested as a solubilizer for the hydrophobic drug Piroxicam in PBS buffer. Different copolymers were synthesised with and without the presence of Piroxicam in their reaction mixture. After precipitation, which was able to remove around half of the original Piroxicam, the maximum solubility was determined via UV-Vis. The “imprinted” polymers performed better than the polymers synthesised without Piroxicam, reaching a maximum solubility of 2.5 $mg \cdot mL^{-1}$.

Compared to other studies that investigate different solubilizers, the specific 1:2:1 BMA:DMAEMA:MMA copolymer performed well. Keeping in mind that Piroxicam is a proof on concept to proof that “imprinted” polymers have a specific sequence that is relatively easy to synthesise compared to single monomer insertion strategies. A faster synthesis opens up possibilities for a broader range of tests for future applications.

Chapter 3: A Predictive Machine-Learning Model for Propagation Rate Coefficients in Radical Polymerization

1. Introduction

The correct assessment of reaction kinetics and the determination of reliable rate coefficients for reactions are often tedious and require sophisticated methods. This is especially true for kinetic rate coefficients in radical polymerization. In polymerizations, to make it more complicated, rate parameters do not merely predict the rate of a polymerization. They also play a crucial role in the design and synthesis of novel materials since individual reaction rates influence the structure. A meaningful prediction of monomer conversions, molecular weights, and polymer dispersities is only achievable if the reactivity information of the monomers can be correlated with the rate coefficient of chain propagation and termination at minimum. The invention of the pulsed laser polymerization - size exclusion chromatography (PLP-SEC) method 35 years ago marked a turning point in investigations in polymerization kinetics by providing highly reliable measurements of propagation rate coefficients in a comparatively simple fashion.

In 1996, Lyons et al. published their findings about how intramolecular transfer does not occur when using high pulse frequencies in their PLP-SEC measurements.[68] Before, these intramolecular transfer reactions posed issues with branched polymer chains and led to inaccurate measurements. Hence, older literature or literature that does not make use of high pulse frequencies have a less accurate value of $K_{p.}$ [69, 70, 71] Over the years, several monomers have been investigated by this technique, and International Union of Pure and Applied Chemistry (IUPAC) working groups have benchmarked data for a number of important monomers. PLP-SEC allows for determinations with relatively high precision – typically an error of 10 to 20 % is estimated. Yet, no unifying approach exists to date that correlates the structure of a monomer with its rate of propagation, and therefore no meaningful prediction of kinetic data can be made. Some trends are known in specific families of monomers, most notably the (meth)acrylates. For example, the – at first glance counterintuitive – increase of the propagation rate coefficient (k_p) with the length of the ester side chain.[72, 73] The same correlation between the chain length of the ester in linear (meth)acrylates and the k_p value is confirmed by Ballard et al.[74] Yet, already smaller differences in the structure such as branching vs linear ester chains are not captured in the literature.[75] The traditional way to predict k_p is to use high level ab-initio quantum chemical calculation. While by themselves highly interesting, these calculations have for some monomers confirmed experimental values, but they struggle to make absolute predictions.

It is largely known that the propagation rate coefficients of monomers depend on a series of factors. One of the hypotheses is that molecular weight is one of these factors. Since linear (meth)acrylates are only a fraction of the existing monomers, the choice was made to use molecular weight instead of the number of carbons in the ester chain. Since the molecular weight and the ester chain length are directly proportional for linear (methacrylates), the molecular weight was considered an adequate

substitute while still able to get a consistent value for the other parameters. Another factor that is believed to have an influence is the resonance stability of the propagating radical. Depending on the substituent on the carbon that contains the radical, the dissociation energy (or radical stabilizing energy) varies.[76] The more stable a radical is, the better the transfer reaction occurs.[77] H-bonding has been identified to cause significant rate effects and polarity is speculated to impact k_p. Mavroudakis et al. demonstrates that monomers and solvents which are capable of hydrogen bonding have a significant influence on k_p.[78] One of the examples is the comparison between two structurally similar molecules 2-hydroxyethyl acrylate (HEA) and butyl acrylate (BA). HEA, which is capable of H-bonding, has a 25% higher k_p compared to BA.[78] In general, hydrogen bonding affects the propagation rate and results in higher k_p values. [78] [79] The effect of the solvent on k_p is an interesting study in and on itself but was not considered in this work. Instead, only k_p values whose polymerisation occurred in bulk were considered. For future studies, the solvent effect would be extremely interesting to investigate but it currently falls outside the scope of this project.

These factors have, however, always only been investigated as insular effects, and no general theory could so far be formed that would unify all different aspects in one approach. Part of this issue might be that in the classical approach, physical chemists look for causation rather than a correlation or association. In complex interdependent systems, this can be a difficult endeavour since accurate data is often difficult to find, and actual causations might not be obvious. Association is much simpler to establish though via purely statistical approaches. Machine-learning (ML) harnesses this relative simplicity to predict complex behaviour of systems. Hence, the question could be raised if it is possible to correlate complex propagation rate coefficients with fundamental and readily available information about monomers without the attempt to establish exact equations reflecting the underpinning processes. If such an approach is successful, the resulting correlation can ideally be used to reach a better theoretical understanding. More importantly though, if statistical association is successful (without necessarily identifying the underlying causal mechanism), then rate coefficients would in principle become predictable. Propagation rate coefficients provide an ideal scenario for testing this hypothesis since relatively accurate coefficients are indeed available for a series of molecules. In the following, we discuss if these kinetic rate coefficients can indeed be predicted on a purely statistical basis rather than using high level ab-initio calculations using transition state theory.

2. Data Selection

Any statistical model approach requires a dataset that can be used to train a model. As mentioned above, k_p is ideal since the IUPAC has benchmarked rate coefficients for a series of monomers. k_p data for these monomers can be assumed to be fairly accurate. In fact, recently an online database was established that allows to retrieve these coefficients directly. In order to benchmark a monomer, IUPAC typically requires more than one laboratory to provide data. In addition, several laboratories have provided sole PLP-SEC data that can still be regarded as fairly reliable. We collated data for close to 40 monomers that we deemed reliable (it should be noted that the IUPAC also defined reliability criteria that make such selection possible). We omitted acidic monomers, since it is known that they are primarily governed by pH, and hence are outliers in the complete set of available data. Most data is available for the monomer families of acrylates and methacrylates. It is known that these monomers can be correlated within their respective families, allowing for some inference from one family to the other. All other monomers have no known quantitative correlation, even though it is common knowledge that radical stability plays a major role in predicting their reactivity. A complete list of monomers and their respective values in Arrhenius form are given in the supporting information. For the sake of this work in this study, four distinct groups of monomers have been identified, these being as described above (i) acrylates (n = 13), (ii) methacrylates (n = 16), (iii) monomers exhibiting strong H-bonding effects (n = 5) and (iv) 'others' (n = 7). An alphabetic list of monomers with their abbreviation and SMILES notation can be found in the Appendix (Table S1). For every monomer, the activation energy E_a and pre-exponential factor A were collected (Table S2). Using the Arrhenius equation, the natural logarithm of the rate constant $\ln(k_p)$ was calculated for four different temperatures.

For every single monomer, different descriptive parameters were researched and noted down (Table S3 and Table S4). Most of the initial parameters, such as the molecular weight and the Gibbs Free Energy, were generated using ChemDraw and thus available for every single monomer. However, later parameters such as the dissociation constant or the the experimental boiling point are unavailable for all the entries. Unfortunately, a monomer needs to have all the desired parameters available or it cannot be used in the regression. There are three main selections of data used in the future: (1) all available data, (2) available data with the dissociation constant and (3) available data with the ChemSpider parameters. An overview of how many entries are in each category with each selection is displayed in Figure 3-1.

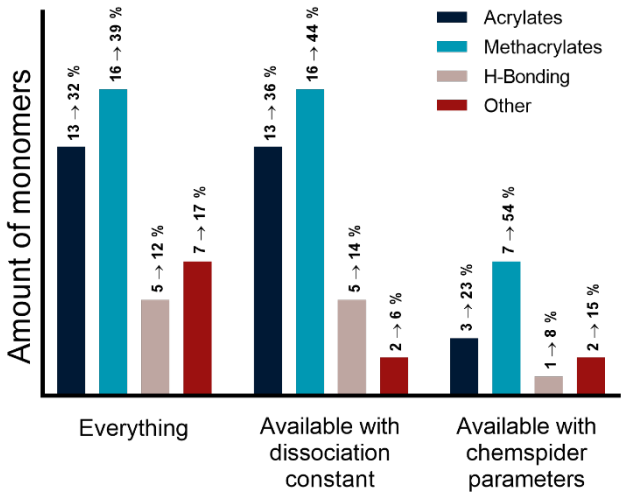


Figure 3-1: : Bar plot of the number of monomers and the respective percentage in the corresponding classification group: acrylates (blue), methacrylates (red), H-bonding monomers (green) and others (black).

3. Methodology

Data was fitted via multivariate linear regression, which is a combination of multiple linear regressions on independent variables for one dependent variable. A popular variant of least-squares regression to avoid overfitting is the Ridge regression method. Ridge regression considers that various features might display collinearity. In this instance, the ordinary least squares method is modified to minimise the squared absolute sum of the coefficients, known as so-called L2 regularisation. This is done by introducing a tuning parameter, lambda (λ), in the cost function (see supporting information). Take note due to Python syntax, the lambda term has to be configured via the “alpha” argument. Similar to ridge regressions, LASSO regression can be employed. In Lasso regression, the cost function minimises the absolute sum of coefficients instead, known as L1 regularisation. As a result, Lasso regression assists to reduce overfitting and can be used in feature selection and hence for physical interpretation posteriori. λ is a scalar variant, and thus requires continuous data to be standardised for both ridge and lasso regression. When λ increases, the bias increases and the variance decreases. Bias hereby describes how well a model matches a training set, whereas variance describes how much a model changes when it is trained with a different training set. High bias infers a weak match with the training set, while low bias indicates a very close match. High variance means that a model is flexible and prone to overfitting. Low variance means that a model is robust and will not change should any part of the training set be altered. Thus, low bias and low variance are generally preferred. For an unbiased estimate of the model performance, the final model fit would use an average of all the final predictions of each monomer. In this way, bias is minimised because all the data is being used for training the model (n-1) times, and variance is minimised since all the data is being used in the validation set once. As a result, the Leave One Out Cross Validation (LOOCV) allows to compare predictions of the target value as a function of its experimental value. In the following we will always use this representation to discuss the results obtained (Figure 3-2).

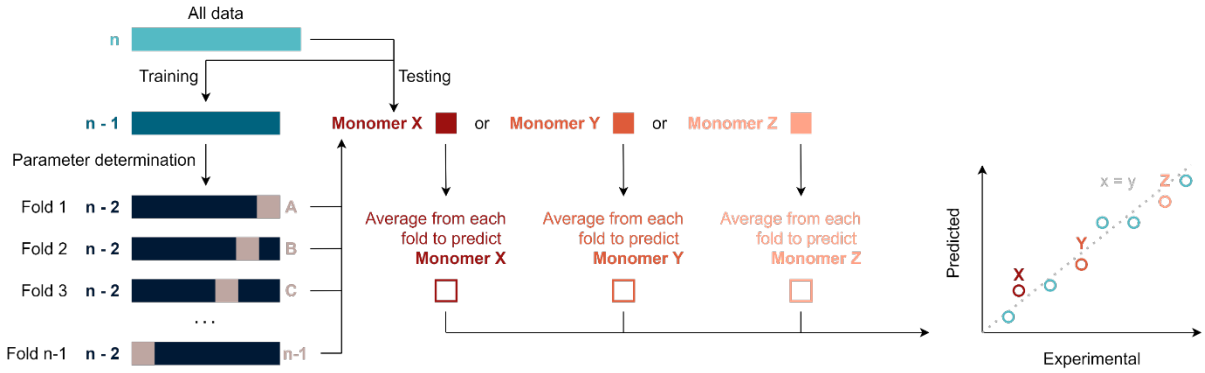


Figure 3-2: Overview of how the data is split for the cross validation using a Leave One Out Cross Validation (LOOCV).

When considering monomers in radical polymerization, several molecular properties, in the following referred to as 'features' in line with data science terminology, are obvious to consider. It is known from literature that polarity is an important quantity that has direct influence on propagation kinetics. Thus, the dipolar moment of the monomer is of general interest. Note that experimental data is mostly available for bulk polymerization, hence where the polarity of the monomer concomitantly influences radical reactivity and the solvent environment. Further, the length of ester side chains in (meth)acrylates are known to at least indirectly correlate with k_p. Thus, molecular weight was added as a further feature. Already when collating experimental data on dipolar moments, it is unfortunately evident that gathering such data is by far not trivial, and generally leads to scattered datasets. To solve this issue, we decided to include calculated data as features in our analysis. To this end, we used the General Atomic and Molecular Electronic Structure System software package (GAMESS, version: 2018, R1).[i] interfaced with the software ChemDraw 3D, and also used data provided by ChemSpider, and the ACD/Labs Percepta Platform - PhysChem Module predictions listed therein. Via GAMESS, we accessed dipolar moments, boiling points, melting points and Gibbs free energies for each monomer under investigation (relative to ethylene as the simplest radically polymerizable monomer possible), using a low-level HF calculation method. ChemSpider provided some experimental data on boiling points and refractive index, and predictions for the same, plus predictions for surface tension and polarizability. Comparison of calculations with available experimental values showed that the theoretical values are certainly not perfect when examining absolute values, but are reasonable when comparing series of monomers with each other.

4. Algorithm testing and parameter selection

In initial testing, all the different regressions (linear, ridge and lasso) are compared to each other using both the k_p and the $\ln(k_p)$ as the predicted value (Figure 3-3). It is clear that just using the k_p provides negative predictions which makes the $\ln(k_p)$ a more desirable parameter. Furthermore, one can see that the predicted values for the k_p of the methacrylates are very clustered. The performances of the Ridge regression and the Lasso regression were very similar but it was opted to only display the Ridge regression in future analysis since its r^2 value was often slightly better.

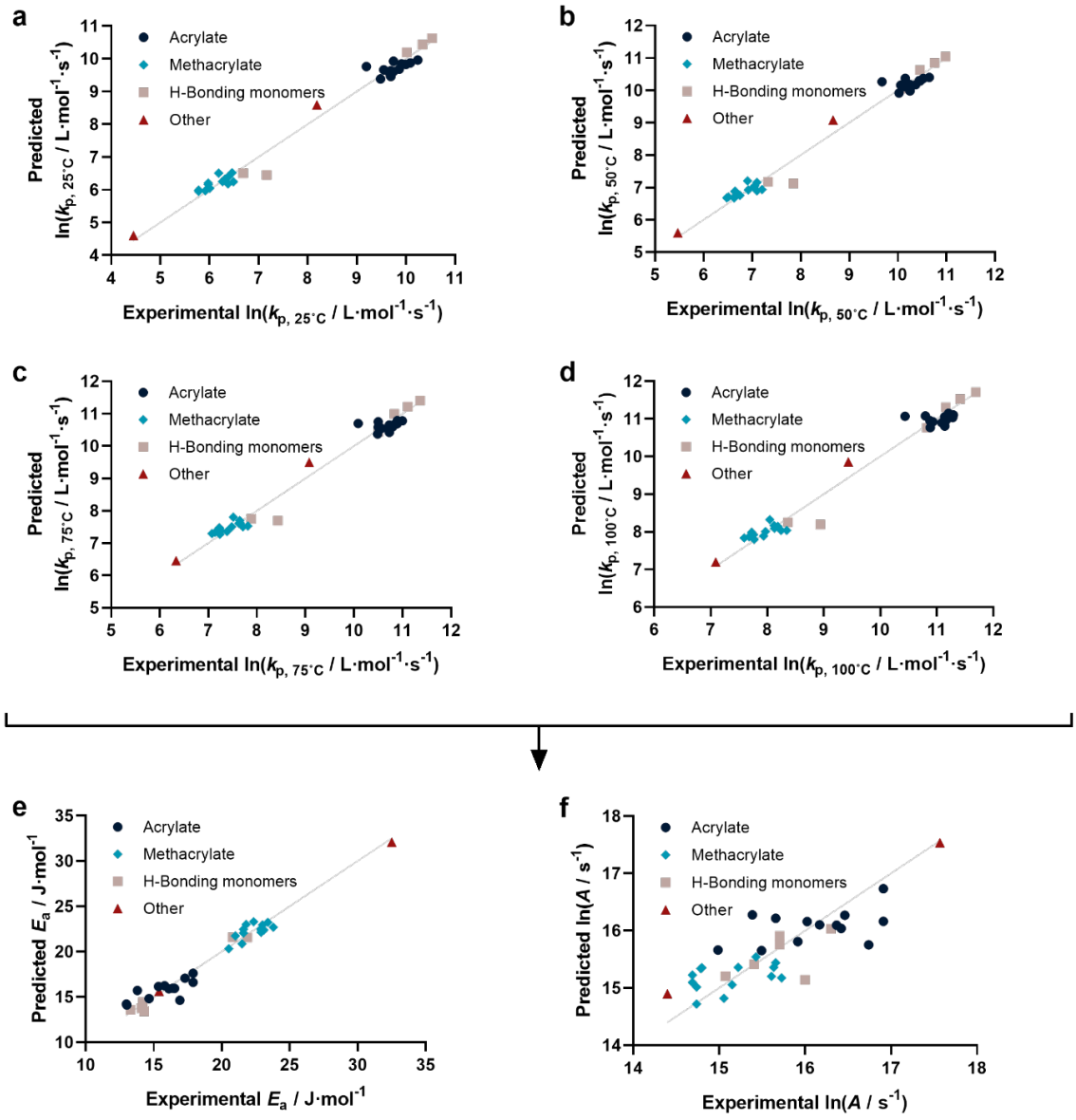


Figure 3-3: Comparison of different regression models for both k_p and the natural logarithm of k_p.

Before describing the regressions on the complete feature set, it is worthwhile to examine the state-of-the-art in predicting propagation rate coefficients. Two influences are known with fairly high accuracy, that is that the propagation rate coefficient increases with the length of the ester side chain in (meth)acrylates; and acrylates propagate up to a factor 100 faster than methacrylates. For all other monomers, no clear correlation has to date been quantified. Thus, in principle, for acrylates and methacrylates individually linear regressions with molecular weight can be carried out. Based on literature assumptions, this should yield some predictivity. Indeed, when plotting experimental k_p as a function of molecular weight, a slight tendency towards increasing molecular weight can be observed. The overall correlation is, however, less than satisfactory. This is due to the list of monomers containing examples that have branched side chains, or that are associated with

H-bonding. Nonetheless, this simplistic model can be used to derive a more general visualization of data. Using the LOOCV method, individual k_p was predicted for each acrylate and methacrylate. For this, we fitted each group of monomers individually as would classically be done. This results in a predicted value for each of the monomers. This residual, predicted value is then plotted against its measured value.

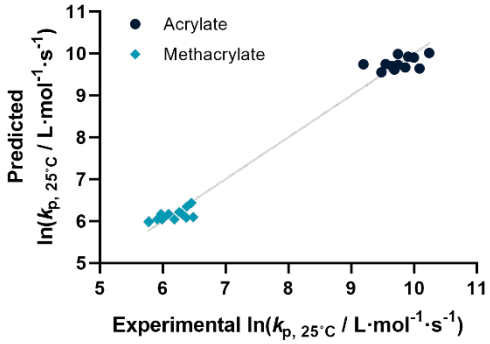


Figure 3-4: The predicted versus experimental values of the ln(kp) at 25˚C for a dataset n=29 using the leave one out cross validation on a linear regression. Predictions are determined using a separation between acrylates (blue) and methacrylates (red)

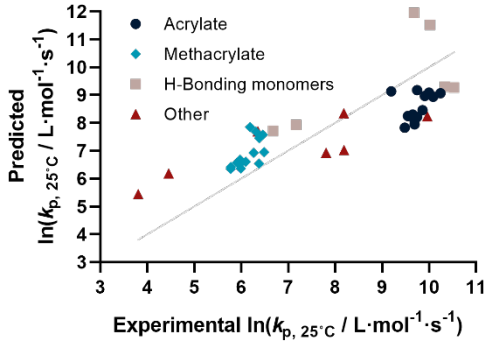


Figure 3-5: The predicted versus experimental values of the ln(k_p) at 25˚C for a dataset n=41 using the leave one out cross validation on a linear regression. Predictions are determined using the molecular weight (Mr), dipole moment (DP), boiling point in Kelvin (BPK), melting point in kelvin (MPK), and Gibbs Free Energy (GFE). (r^2 = 0.426)

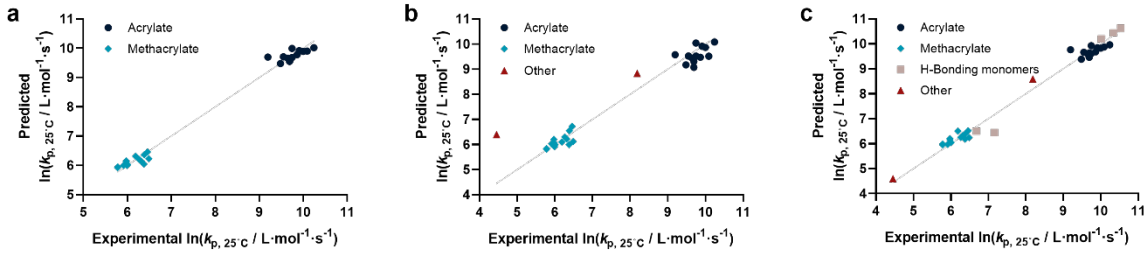


Figure 3-6: a) Comparison of predicted vs. experimental ln(kp) for isolated acrylates and methacrylates (r2 = 0.991), b) after inclusion of non-meth(acrylate) monomers (r2 = 0.940) and c) all monomers using available data (r2 = 0.985).

Ideally, if the experiment was error free and if the prediction was 100% accurate, a linear plot with a slope of one should be observed. The sum of squares of this fit can be used to quantify the predictivity of the underlying model used. The outcome of this data procedure is shown in Figure 3-4. It should be noted that for all predictions, the logarithm of k_p was used rather than its actual value; this prevents the model from predicting non-physical negative values. As can be seen from the figure, the linear regression representing the state of the art basically predicts two plateaus, one for methacrylates and one for acrylates. Rather than predicting an increasing k_p with molecular weight, it thus results in not more than a rough average per monomer group. This shows that the perceived correlation shown in Figure 3-5is not statistically robust (while not necessarily wrong), it cannot be used to predict any unknown monomer k_p. This is, taking the large scatter of data into account, not really surprising.

Coming from this rather sobering result, we then extended the feature set of the regressions, taking all calculated values into account, hence molecular weight, dipolar moment, melting and boiling point and Gibbs free energy values. The resulting prediction is almost random, as can be seen in Figure 3-5. No obvious correlation can be seen. Hence, we moved forward by adding further information to the feature set. To achieve this aim, we first broke the number of datapoints down, and isolated the (meth)acrylates from the list. Then, we introduced a binary differentiation for acrylates (one) and methacrylates (zero). This alone leads to a reasonable representation of data, showing that the calculated physical properties of monomers have a positive effect, and are aiding in the prediction of rate coefficients. However, several difficulties remained. For example, hydroxy ethyl acrylate was not appropriately predicted. Also branched monomers, while improved compared to the state-of-the-art representation in Figure 3-4, still showed significant deviations. To solve these issues, the feature set was extended by two parameters. One parameter described the inductive effect of the side chain, the other quantifies H-bonding between monomers. For both effects, literature was screened, and property tables provided by high level quantum calculation could be used. It should be hereby noted that H-bonding is not easy to quantify, and only the presence of major functional groups was accounted for. Despite the shortcomings of this process, a very good prediction is obtained in this way. Figure 3-6a depicts the case fitting of acrylates and methacrylates without inclusion of strongly H-bonding monomers. It should be noted that both LASSO and Ridge regressions yield similarly reasonable results, while conventional linear regressions perform significantly less well. Nonetheless, the r^2 value of the plot shown in Figure 6a is 0.991, underpinning its overall high quality. Statistical analysis of the LASSO regression shows that either the melting point or the boiling point can be used, both information is not required since they display high collinearity. The same was true for surface tension data and refractive index data. Both features were practically redundant when polarizability was used. All other features do contribute to the results.

Yet, when including all monomers back into the fit, also this procedure still yielded an unsatisfactory result. The obvious reason for this is that none of the features sufficiently describes resonance effects, which play a major role in reactivity of monomers. Surprisingly, while qualitative orders of reactivity are obviously known for practically all polymerizable vinyl monomers, no quantitative data on the resonance stability, or radical stability could be found in literature. The closest information that we identified were dissociation constants of macroalkoxyamines, that were determined for nitroxide-mediated polymerization from Electron paramagnetic resonance (EPR) spectroscopy. At least for styrene, acrylonitrile, as well as for an average of methacrylates and acrylates, numeric values could be assigned. We normalized these dissociation constants, and provided our best prediction, shown in Figure 3-6b. Finally, since this plot yielded a very reasonable fit, we then directly also included the H-bonding monomers, as depicted in Figure 3-6c. Again, it is an interesting observation, that correlation of hydrogen (H)-bonding has also a positive predictive effect for example for styrene (the datapoint with the lowest overall k_p). As can be seen, both acrylonitrile and styrene (black triangles) fall very well on the line with the (meth)acrylates. r^2 in this case is 0.986, which is a very good result considering that both added monomers have nothing in common with the other monomers in question regarding all other used features. It can be assumed that if actual data for resonance stability of propagating radicals become available for other monomers (such as vinyl acetate for example), that also these monomers can then be adequately correlated. Such data can probably also be derived by quantum chemical calculations, and with that also removing any need for experimental data in the current approach. Overall, the ridge regression analysis for weighting of the various features shows that resonance stability by far is the most important feature, overshadowing the previous importance of dipolar moments and substituent effects.

At this point, a good general correlation of all data points had been reached. Closer inspection of sub-datasets revealed, however, a further interesting correlation. While the theoretical values provided by ChemSpider/ACDlabs did generally contribute to the success of the fitting, the accuracy of the fits improved almost by an order of magnitude when only data was fitted for which experimental boiling points were available. The results of the fit for this subset are shown in the appendix (Figure 4-23). The correlation is achieved in this plot has a value r^2 value of 0.953. More interestingly, the r^2 of before the cross validation was close to perfect with a value of 0.999. it is theorised that the decrease of r^2 after cross-validation is largely due to the lack of datapoints. It can, however, be speculated that availability of experimental boiling points improves the accuracy of the prediction of polarizability, which in turn is an important parameter. This hypothesis, can at this point though not be tested since the ACDlabs predictions are provided by ChemSpider, and are not directly accessible to us. One of the theories we tested was to check if one gave an increased importance, a so called weight, to those 13 datapoints in the whole dataset, one could achieve better results. Disappointingly,

there was no massive improvement observed as shown in the appendix figures (Figure 4-28 and Figure 4-29).

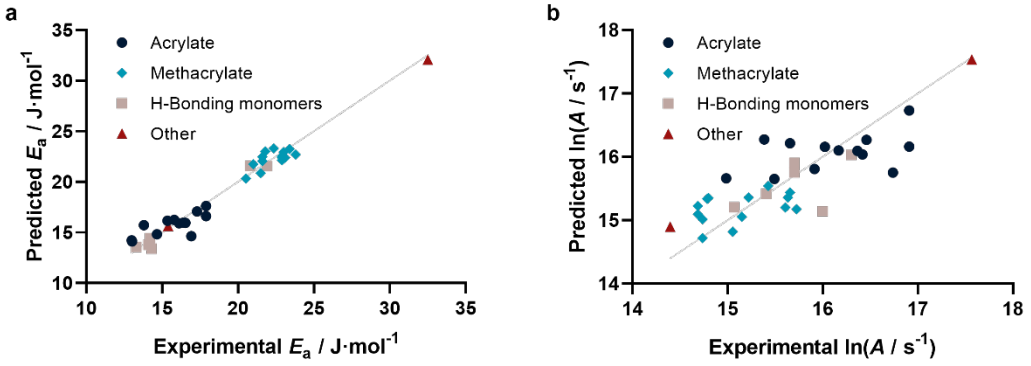


Figure 3-7: a) Activation energies (r^2 = 0.963) and b) Arrhenius factors (r^2 = 0.666) determined from prediction of individual k_p at four different temperatures.

Regardless, with this method at hand, we tested if not only individual k_p can be predicted, but also activation parameters. To this end, we calculated k_p for 25, 50, 75 and 100 °C based on experimentally derived activation energies. As for the 25 °C data in Figure 3-6, similarly reasonable fits were obtained. For each individual monomer the predicted k_p values at the four temperatures were fitted to the Arrhenius equation, yielding a predicted Arrhenius factor A and a predicted activation energy E_a. Using the same representation as for k_p, it can be shown (Figure 3-7a and Figure 3-7b) that also the activation parameters are well predicted by our model. Specifically, E_a is represented very well, while ln(A) shows some more scatter. This scatter is a result of the sensitivity of A on small variations in E_a and correlates with the typically also higher scatter of experimental data for this value. It should be noted that it is in principle also possible to directly correlate E_a and A instead of k_p at a distinct temperature. However, this would require a simultaneous regression of both parameters, since A and E_a are highly interdependent. If fitted individually for E_a, a reasonable direct correlation is found, yet not for A. The approach to determine both values via individual k_p's is hence more successful and reliable.

Until this point, we only demonstrated that the data can be successfully correlated. Of course, the aim of any such investigation is to predict rate coefficients for monomers for which no experimental data is available. With the good correlation demonstrated in Figure 3-6, predictions of rate coefficients should be possible for monomers for which no experimental values of k_p are yet available. Since practically all features in our model are based on calculated features, it is fairly straight forward to include further monomers into the list. Given the almost perfect alignment of Figure 3-6a, best predictions should though be available for monomers that have experimental boiling points known. As monomers in question we chose ethyl acrylate, propyl acrylate, cyclohexyl

acrylate and propyl methacrylate. We used the boiling point subset model to then predict k_p at 25˚C. These monomers are an interesting comparison since it is known that propagation rate coefficients increase generally with the size of the ester side chain for both acrylates and methacrylates, and reference data are available for the corresponding methyl and butyl (meth)acrylates. All ML-predicted k_p values are given in Table 1. The coefficients of a couple of different models were calculates and displayed in the Appendix A (Table A8, Table A9, Table A10 and Table A11). It is interesting to note that the calculation of these coefficients (and their importance, which is the coefficient multiplied by their variance) are interesting but not necessary to obtain predictions since that information is already part of the program.

Table 1. Model predictions of propagation rate coefficients for monomers for which no experimental k_p data is available. Predictions are based on the data shown in Fig 3b

Monomer	k_p(25 C) / L·mol^{-1}·s^{-1}
Ethyl acrylate	9300 < 13212 < 18769
Propyl acrylate	9542 < 13555 < 19256
Cyclohexyl acrylate	10258 < 14573 < 20702
Propyl methacrylate	**272 <** 386 **< 548**

As a first observation of the data in Table 1, one can see that qualitatively the order of predicted propagation rates in the acrylates is correct. A slight increase in rate for the ethyl to the propyl acrylate is seen, while the cyclohexyl acrylate monomer exhibits the highest rate coefficient. For both the acrylate and methacrylate the correct order of magnitude is predicted (which is, given the large number of monomers used in the correlation from these two families not surprising). While the order of monomers is correct, absolute values are not fully matched. Starting from the experimental propagation rate coefficients for methyl and butyl methacrylate, a k_p between 323 and 370 would be expected. The value for propyl methacrylate is very close in the range of expectations. Given that size-exclusion chromatography, which is key to experimental k_p determinations is commonly associated with an error of up to 20%, this match is exceptionally good, and outperforms any prediction based on ab-initio calculations provided so far. For the acrylates, using the same comparison, a k_p between 13130 (methyl acrylate) and 16380 (butyl acrylate) would be expected. Again, the predictions made by our regressions meet this range. For cyclohexyl acrylate, no clear expectation can be given, yet the produced value seems to be a reasonable estimate. It will be interesting to see in the future if these values will be confirmed by experiments. It should be noted that using the full dataset as given in Figure 3-6b and Figure 3-6c yields predictions within a 20-40% deviation from the ones given in Table 1. Given the lower overall r^2 of these fits, this is not surprising, and the data produced based on the smaller dataset in Table 1 should be more appropriate and precise. Regardless, we have made our Python script for the prediction available via the Github platform, so that other researchers can make their own predictions based on need.

5. Results and discussion

An unexpected time-consuming part of this chapter was the collection of data for the database. While supervising the results of the algorithm, certain parameters were added and/or removed. Of the different regressions (linear, Ridge and LASSO), it was observed that the Ridge and LASSO regression performed similarly but that the Ridge regression gave slightly better r^2 values. Hence, most of the plots that are displayed were generated using a Ridge regression. From the parameters, it was concluded that the optimal combination investigated in this study was: (1) the molecular weight**,** (2) a distinction between the type of monomer, (3) the inductive effect of the tail group, (4) the effect of H-donors, (5) the effect of H-acceptors, (6) the dissociation constants and (7) the polarizability. Combinations of different parameters were added manually within the code, so the results could be manually checked (making it a supervised learning model). This way, a more controlled selection of the parameters was obtained. The idea is that with controlled selection of parameters, it would lead to a bigger understanding which parameters are the most important/the most essential in the final algorithm. Overall, the desire is to work towards an accelerated polymer design, an idea that has also been discussed by Patra et al.[80] In this review, they emphasise the importance of moving away from a traditional trial and error design. One of the general challenges (something that was experienced in this work as well) is that in the available dataset there is less experimentally measured data available. One of the big (and disappointingly boring) findings of this work is reflected in other studies as well: namely that there needs to be a bigger, more complete and consistent experimental dataset of synthetic chemicals.[80] Interestingly, these databanks, such as the protein databank, are already available for biomolecules.[81] In this work, there was an attempt made to use experimental data (such as that of the boiling points, polarizability etc.). Unfortunately, it significantly reduces the amount of datapoints in a set, making it far more difficult to obtain precise predictions. Again, to reduce the size of the error margin more experimental data is necessary.

In literature, rather than making use of physical parameters as the monomers descriptors, techniques such as One-Hot Encoding (OHE) are used to describe the structure of the monomer (or any other structure) itself. With OHE, the structure of the monomer itself is changed from categorical data to a numerical data structure (which is easier to use as an input for ML).[80] OHE is also used to translate the SMILES notation into a binary image.[82] Property colouring, such as described by Webb et al., makes use of a featurisation that changes the polymer structure into an image which works well with ML in coarse-grained polymers.[83] Autoencoders on the other hand, map discrete molecular representation to a continuous latent space. Thus, autoencoders can be used to transform the SMILES notation into continuous variables.[84, 85] Motif-based fingerprinting, used for representing small molecules, is based on the quantity of specific atoms present in a molecule and the chemical bond formation/coordination with other atoms in the same structure.[86]

The paper of Shi et al. demonstrates a Quantitative Structure-Property Relationship (QSPR) model where they predict the k_p value by using the temperature and a Norm-Index (*NI*) descriptor.[87] The two key differences are that (1) they have multiple entries of the same monomer for different temperatures (rather than basing the entire model on just the E_a and A values) and (2) they have one vector based descriptor rather than multiple parameters as descriptor. The first difference has the advantage that the database is much bigger, but it also means that some monomers are more represented than other ones. The second difference is a quite elegant solution about how they, in essence, "describe" the monomer to the model. It would be extremely interesting to see if combining this *NI* descriptor with some of our parameters would enhance the predictive results.

Overall, the difference with using techniques that gather information directly from the structure of the molecule is that it gives us less information in which physical parameters play a big influence in the polymerisation. Admittedly, the method described in the chapter is far from perfect but it does provide unique information (/confirmation) how polymerisations effectively work. Ideally, when a ML algorithm is coupled to an automated experimental setup, it would be interesting to include the solvent effects on the polymerisation.

6. Conclusion

The predictions made based on the model presented herein may not be able give a 100% accurate representation of small effects such as ester side chain length, yet they provide very reasonable data that can be used for future work. Accuracies are within standard error limits of experimental determinations, and relative effects are accurately predicted. No prediction of k_p values with the same accuracy by any other theoretical method has to date been provided. With increasing data coming available from experiments, and potentially with further refinement of feature calculations, we hope the accuracy of our method can be further refined in future work. The match of data currently seen for monomers with available experimental data on boiling points should be extendable to the full dataset. Generally, though, it is remarkable that only very little experimental data input is required in all predictions, whereby experimental boiling points seem to play only a role in refinement of other theoretical data. Regardless, even though this hypothesis will still need experimental validation, it seems that simple boiling point measurements in conjunction with radical stability data, may be sufficient in the future to predict accurate propagation rate coefficients directly for any kind of polymerizable vinyl monomer.

Regardless of the predictivity of the regressions, this work presents for the first time a full correlation of complex kinetic rate coefficients with structures. Specifically, no model existed to date that was able to correlate monomer structures with propagation rate coefficients in radical polymerization, and already the satisfying match of all monomers as seen in Figure 3-6c is already a large gain in knowledge, even in its simple form of sets of regressions, as used in here, has tremendous potential for simple, and readily available predictions of kinetic data. It should be noted hereby that the performed predictions occur within seconds, and can be easily implemented in kinetic modelling codes in the future.

Chapter 4: An Exploration towards Predictive Machine-Learning Models for Reactivity Ratios in Radical Copolymerization

1. Introduction

When designing a polymer for a specific application, there are more possibilities when using copolymers then homopolymers. Specific tuning of ratio of monomers build in the polymers, the sequence of the monomers and the structure of the overall polymer allows for a very broad amount of possibilities even when you start with the same starting materials. A popular review from Adams et al. about amphiphilic block copolymers for drug delivery, currently almost twenty years old, has been cited by hundreds of other people.[88] These application range from cancer therapy (Känkänen et al.), functional antibody delivery (Koch et al.), temperature sensitive drug carriers (Dou et al.) and various more.[89-91] In the correct circumstances, the polymeric micelles made from amphiphilic block copolymers can serve as a carrier for hydrophobic drugs. On top of that, it is possible to tune the shape of the micelles by varying the mixing parameters when performing the self-assembly, as demonstrated by Buckinx et al.[11] Even random copolymers have their uses in drug delivery and tissue engineering applications.[92] Nevertheless, especially in biomedical applications, it is essential to have the final properties exactly right. Polymers (and specifically copolymers) have great potential in biomimetics, artificial peptides, molecular recognition or data storage. Some examples include self-assembling amphiphilic block co-polymers as a use in biomimetic interfaces from Avsar et al. or channel-based membranes from Lang et al. [93, 94] Even more interesting is using nature as a template such as in the research of Salas-Ambrosio et al. where they use synthetic polypeptide polymers as analogues of antimicrobial peptides.[95] One can even combine the “natural” peptides and the “synthetic” polymers to create peptide/protein-polymer conjugates as outlined by Gauthier et al.[96] Even already in the nineties, Seymour et al. described synthetic polymers that are conjugated to monoclonal antibodies.[97] An example of polymer sequences as a data storage application is highlighted by Min Lee et al. where they use defined aperiodic sequences of monodisperse copolyester or Kaempf et al. where they use polymers as substrates for data storage.[98, 99] Vrijsen et al. describes the use of molecular weight distribution fingerprints as a way to generate secure data encryption.[100]

Hence, the ample application possibilities make the polymer chemistry field interesting for many different industries. Currently, there is much interest regarding sequence controlled- and sequence defined polymers using reversible-deactivation radical polymerisation (RDRP), formally known as controlled radical polymerisation (CRP). [29] By controlling the precise molecular weight, dispersity and the specific composition of a polymer, one can control its final properties. Implementing the findings (both properties and characteristics) of RDRP polymers in a dataset has two advantages. Firstly, a reliable (and/or benchmarked) dataset serves as an excellent referral for future projects and experiments. Secondly, the dataset can be used to look up existing data, but properties and characteristics can also be used as an input to generate virtual data/predictions via artificial intelligence. However, to predict polymer properties from scratch, one must have access to a vast

amount of data and the experiments to characterise one specific polymer sequence are often time-consuming. Thus, a good start to begin generating this database is by starting small and expanding step by step. Instead of immediately characterising and implementing precise sequences, a better strategy is to cover more general data first. Chronologically, the experiments for this chapter were executed first. But as tempting as it was to immediately start with copolymers, one realised that it was beneficial to focus on homopolymers first, hence Chapter 3.

Copolymerisation chemistry is a field that has lost some academic interest in favour of RDRP over the years. Similar to creating a database with sequence controlled and/or sequence defined polymers, a database with information regarding copolymerisation parameters will have quite some use. One of the more used copolymerisation parameters are the reactivity ratios r_1 and r_2, which Mayo and Lewis introduced in 1944.[26] The reactivity ratios describe the rate monomer "a" will be added to a propagating chain against another monomer "b". The reactivity ratios (r_n) can be defined using the rate constants (k) where rn is equal to knn divided by k_{nm} with n = 1 or 2; m = 1 or 2; and n ≠ m (eq 2) (Equation 1).

$$k_p = Ae^{\frac{-E_a}{RT}} \qquad \text{Equation 9}$$

$$r_n = \frac{k_{nn}}{k_{nm}} \qquad \text{Equation 10}$$

If n = 1 and m = 2, then r1 relates the propensity of monomer M1 to react with radical M1* over monomer M2 to react with M1*. Depending on the values of r_1 and r_2, the composition of the polymer could be influenced (Figure 4-1). Thus, information about r_1 and r_2 gives knowledge indirectly about their properties since the composition of a polymer directly influences how that polymer behaves.

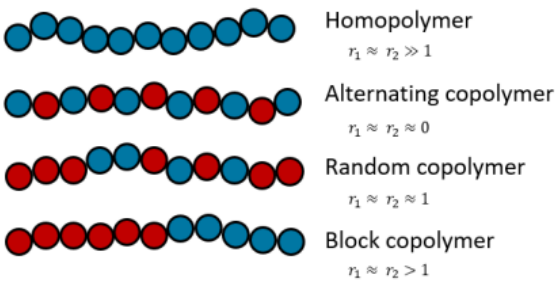


Figure 4-1: Influence of r_1 and r_2 on the final structure

Trying to predict, or estimate, the reactivity ratios is a goal that scientists have researched for decades. Early methods to predict reactivity ratios were by (1) approximating, (2) curve fitting, (3) making use of intersection or (4) linearization.[101] For clarification, *M* is the mole fraction of the monomer in the reaction system and *m* is the mole fraction of the monomer within the polymer.

In the first method, the approximation method, one assumes that the composition of the copolymer is entirely depended on r_1 when there is a low concentration of M_2. The approximated value of r_1 is equal to M_2/m_2. The predictive power is limited, especially when dealing with extreme r values, but the method is straightforward and fast.

In 1944, Alfrey and Goldfinger derived a simple copolymer equation or also known as the copolymer composition equation (Equation 11 and Equation 12). [102] A graph was prepared by plotting the experimental m_2 in the copolymer versus the M_2. Next, a curve based Equation 12 on was plotted and the r_1 and r_2 values were varied until the curve fits the experimental data. There are some disadvantages to this method including bias, extensive calculations and dependence on measuring the weight accurately.

$$\frac{dM_1}{dM_2} = \frac{[M_1(r_1M_1 + M_2)]}{[M_2(r_2M_2 + M_1)]} \quad \text{Equation 11}$$

$$\frac{m_1}{m_2} = \frac{(r_1M_1^2 + M_1M_2)}{(r_2M_2^2 + M_1M_2)} \quad \text{Equation 12}$$

Equation 12 was rearranged to Equation 14, more well known as the Mayo and Lewis equation. This change still poses the same limitations as the previous method since both methods work via the same principle. The difference with the curve fitting method is that one treats the ($m_1M_2^2$/ $m_2M_1^2$) and the (M_2/M_1)[(m_1/m_2)-1] factors as the slope and the intercept of a straight line.

$$r_1 = r_2\left(\frac{m_1M_2^2}{m_2M_1^2}\right) + \left(\frac{M_2}{M_1}\right)\left[\left(\frac{m_1}{m_2}\right) - 1\right] \quad \text{Equation 13}$$

A widespread model to determine the r_1 and r_2 in radical polymerisations is known as the terminal model, resulting in the Mayo-Lewis equation (Equation 3 and Equation 4). In Equation. 3, the r_1 and r_2 are related to the monomer concentrations ([M_n] with n = 1 or 2). This equation can be rewritten in terms of mole fractions of the monomers in the feed (f_n with n = 1 or 2) and mole fractions of the monomer within the polymer (F_n with n = 1 or 2), resulting in Equation 4.

$$\frac{d[M_1]}{d[M_2]} = \frac{[M_1]([M_1] + r_1[M_2])}{[M_2]([M_1] + r_2[M_2])} \quad \text{Equation 14}$$

$$F_1 = 1 - F_2 = \frac{r_1f_1^2 + f_1f_2}{r_1f_1^2 + 2f_1f_2 + r_2f_2^2} \quad \text{Equation 15}$$

The simplicity of the terminal model also comes with limitations. It cannot predict low molecular weight polymers since the equation considers a general steady-state. Furthermore, it does not take the reaction conditions into account. An expansion of the terminal model, the penultimate model, takes the second last monomer into account as well (Figure 4-2). The penultimate model is much more accurate but is more difficult to determine, hence resulting in limited available data.

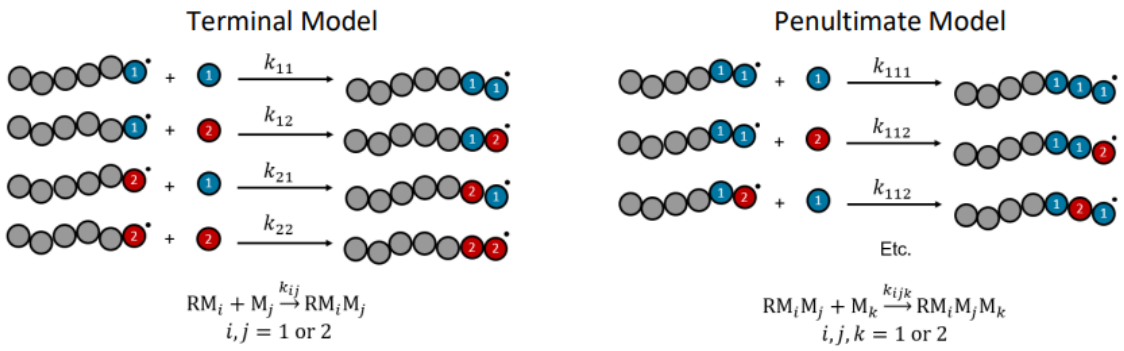


Figure 4-2: Comparison between the terminal model and the penultimate model

To predict the r_1 and r_2 values, the Q-e scheme was developed in 1947 by Price and Alfrey and is still popular because of its simplicity (Equation 16 and Equation 17). *P* and *Q* are the general reactivity of radical *i* and monomer *j*, and e is proportional to residual charges in the reacting groups (Equation 12). The reactivity ratios can be isolated by substituting the four possible outcomes when a polymer undergoes propagation (Equation 6).

$$k_{ij} = P_i Q_j \exp(-e_i e_j) \qquad \text{Equation 16}$$

$$r_1 = \frac{k_{11}}{k_{12}} = \frac{Q_1}{Q_2} \exp[-e_1(e_1 - e_2)] \qquad \text{Equation 17}$$

However, these predictions have severe limitations and are often incorrect. First, the selection of *e* to a given monomer is arbitrary and cannot be changed once set. Therefore, this model is inconsistent once new monomers are introduced. Second, *P* and Q are dependent on the radical to which the monomer is being bound. Third, the steric hindrance, which the penultimate model highlights as a factor in reactivity rates, has not been considered.

In 1950, Fineman and Ross applied a different method and rearranged Equation 12 to Equation 18.[103] This method however, was never as popular as the previously mentioned ones.[101]

$$\frac{M_1(m_2 - m_1)}{M_2 m_1} = \left(-\frac{m_2 M_1^2}{m_1 M_2^2}\right) r_1 + r_2 \qquad \text{Equation 18}$$

Tidwell et al. published in 1965 an improved method of calculating the reactivity ratios.[101] They described how using a non-linear least squared method provided a superior prediction then earlier techniques. Instead of visually deciding which r_1 and r_2 are best (as with the curve fitting method) they minimize the sum of squares between the observed and computed polymer compositions. This method removes the bias of the observer.

Towards the end of the nineties, more advanced machine learning gained interest. Ni et al. implemented artificial neural networks (ANN) to predict copolymer composition drift.[104] Because the

reactivity ratios changed with conversion, they implemented an ANN to correctly map the non-linear relationship of a free radical copolymerization.

In 2005, Rintoul et al. investigated a predictive method based on a case study with acrylamide/acrylic acid mixtures. This prediction method was analytical rather than theoretical and specified in cases of variable monomer reactivity.[105] In 2013, Tan et al. developed Quantitively Structure-Activity Relationship (QSAR) models to predict two parameters *u* and *v*. The so called UV-scheme performed better than the classical Q-e scheme.[106] However, one of the conditions for a successful fit is that their needs to be a high correlation between the experimental and the calculated *e* values. In other words, their *u* and *v* reactivity parameters are only acceptable when their correlation coefficient is greater than 0.9. In 2014, Shrinivas et al. tested a new model called “genetic programming”, a model based on decision trees, that performed a symbolic regression.[107] In their findings, the genetic programming performed better than the classic Alfrey-Price method and artificial neural networks. However, one of the limitations of their model was in extrapolation situations (which could be solved with more data). Also in 2014, Kazemi et al. published an error-in-variables-model (EVM) for the optimal estimation of reactivity ratios in copolymer systems.[108] Fazakas-Anca et al. published in 2021 their work on the Stochastic Gradient Decent Optimization algorithm.[109] In here, they proposed the use of the Fisher criterion as an indicator of the quality of the reactivity ratios.

The main hypothesis was that, without the need of extensive computational power, a predictive tool could be developed that, in future research, could be part of an efficient lab cascade.

2. Data selection

Instead of using the Q-e scheme, it would be interesting to predict the r_1 and r_2 values using a predictive artificial intelligence algorithm. Hence, a database gathered from literature was created. The database was adapted from resources, (see Appendix B), and reliable data was selected. An extra hassle regarding the r_1 and r_2 values is that not all the data points found in literature are reliable, so each data had to be evaluated and then selected. The Mayo-Lewis model hypothesises that instantaneous mole fractions would describe the next step of a copolymer's propagation (Equation 12). This model is often only used early in the conversion processes of monomers to copolymer (which is not always the case). On top of the r_1 and r_2 values, chemical properties of the monomers (such as the SMILES notation, melting point and boiling point) were also determined. The resulting table can be found in Appendix B.

3. Combining predictive algorithms with copolymerization parameter database

In this section, the use of predictive algorithms is explored. Strieth-Kalthof et al. describe how machine learning (ML) modernise synthetic chemistry.[44] One of these predictive algorithms is called a decision tree which is, in essence, a sequence of binary decisions. A schematic overview of the working of a decision tree is taken from Strieth-Kalthof et al., illustrated in Figure 3. For simple datasets or problems, a decision tree can be deduced manually by "asking the right questions". However, ML can generate the “questions” asked at each node for you if it has relevant training data. Since generating these decision trees is entirely random, using an ensemble of multiple decision trees (also known as a random forest) is often more correct. Single trees are prone to overfitting, which is smoothed out by taking an overall average. Because of their relative simplicity, random forests are a good baseline model regarding ML problems in chemistry.

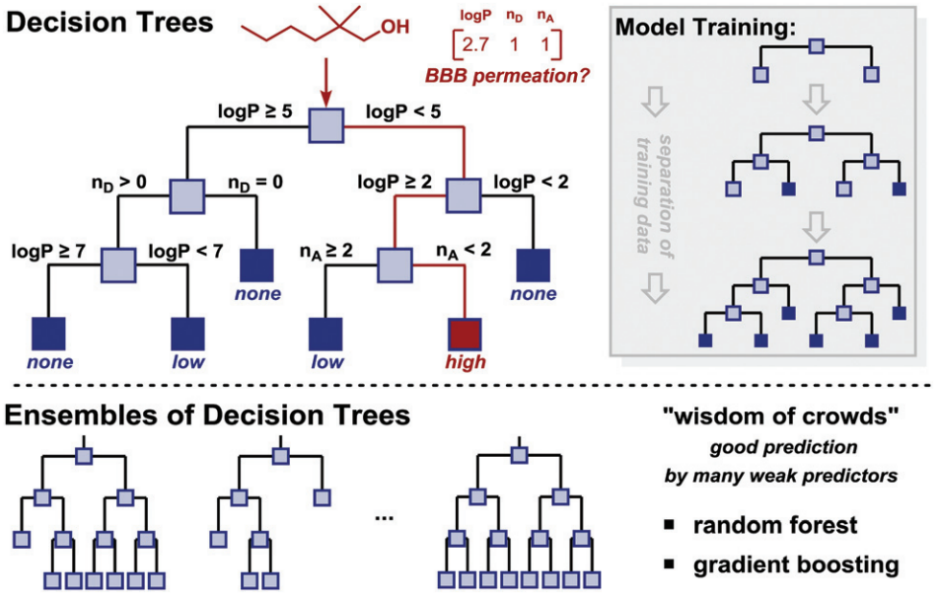


Figure 4-3: Schematic overview of decision-tree-based ML methods, exemplified for predicting the blood brain barrier (BBB) permeation. nD/nA = number of hydrogen bond donors/acceptors. Taken from Strieth-Kalthoff et al.

Using the data of Appendix B and Python, a random forest was generated. The dataset was randomly split into training (90%) and testing (10%) data. When the algorithm generates a decision tree, it will train itself on 90% of the data, knowing the features and labels it expects. When the training is complete, it will then test itself on the test set to determine how well the trained model predicts the value of r1. An advantage of random forests is that compared to a neural network, one can illustrate the ML output (Figure 4-3). Here, one can observe which "questions" the algorithm generated.

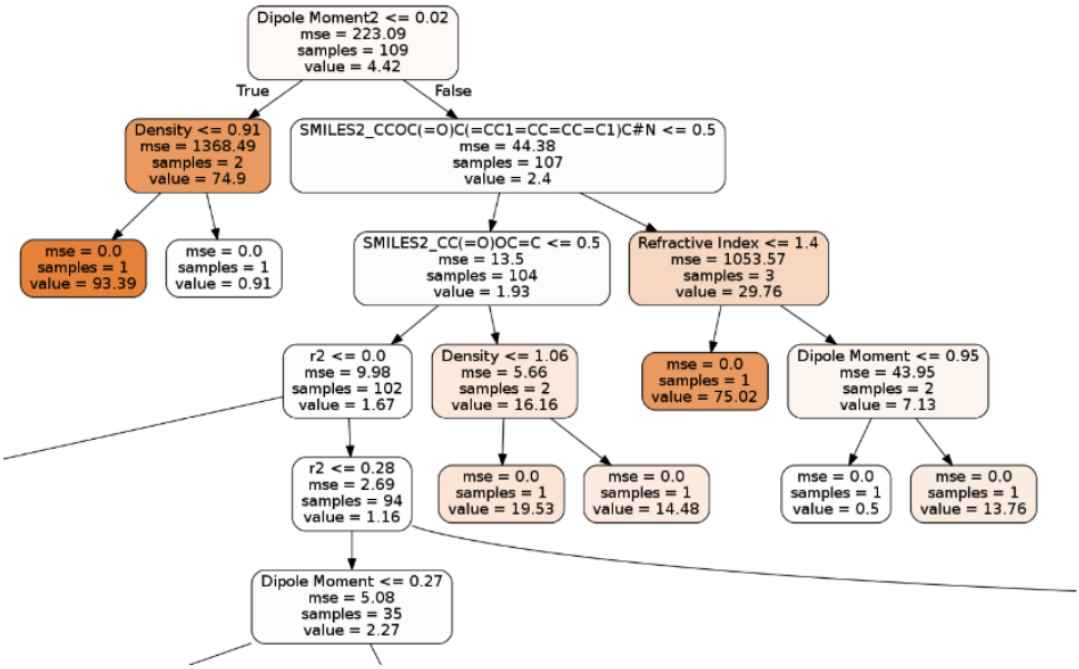


Figure 4-4: Part of the decision tree

Following training on a decision tree (Random Forest Regressor with the number of decision trees equal to 75000), the predictions were compared to the Q-e scheme. The comparison noted was the percentage error predicting each model compared to the actual label value (Figure 4-5). The four types of data shown in Figure 4-5 are as follows: "error without r2" indicates that the decision tree was trained without knowing the value of r2 in the copolymer. It was only trained on the features of each monomer. The added term "min_samples_per_leaf" is a hyperparameter when training the decision tree. If the minimum number of samples is not satisfied to decide the next step, then a different Boolean, or two-element function (yes/no, true/false, etc.), will be considered.

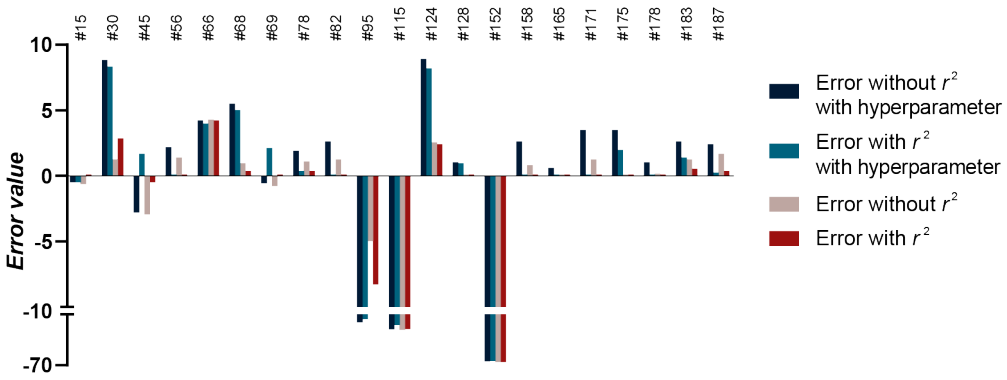


Figure 4-5: Error of test copolymer predictions of r1 when passed through the Decision Tree Network. (n = 75,000), data used Appendix B

As shown in Figure 4-5, if the decision tree can access r_2 during the training process and is not limited by the number of samples required to decide, then the mean absolute error will be minimised.

The mean average error is close to 8 for predictions that did not include r_2 and were limited in samples in the training and predictions. On the contrary, the training and predictions that did include r_2 and were not limited by samples to leaf showed 31.5% improvement, with a mean average error of 5.46.

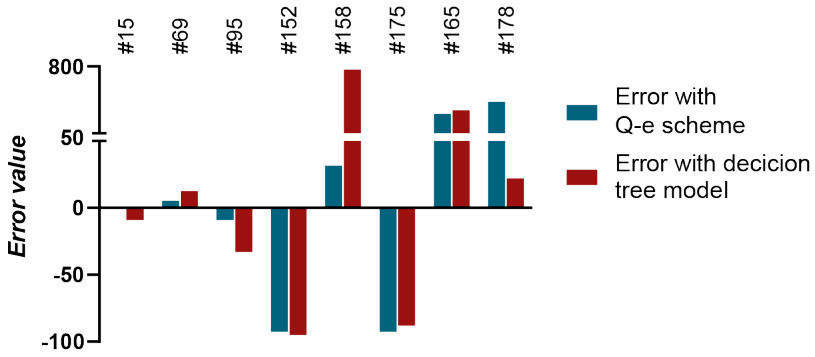


Figure 4-6: Percentage error of selected copolymer predictions of r_1 for the Q-e model and the decision tree network

Overall, 42% of the predicted values included r_2 in the prediction (Figure 4-11) managed to have a predictive error of less than one percent. Interestingly, labels with relatively high values for r_1 (ID's 95, 115, and 152) all had a consistently high error. These relatively large deviations of the prediction to the label were consistent, regardless of whether r_2 was considered in the training and prediction of the decision tree.

A significant target for this study is to provide the framework to improve on the previous models currently used to predict the reactivity ratios. A test set is reserved data where the labels are known but not included in the training. After training on the features and labels, the test set is passed through the trained model. Ideally, a model would have the capability of predicting both r_1 and r_2 when two monomers (and their features) are passed into it. Since there was significant complexity due to many missing data points, the scope of this study was narrowed to only predicting one of the two reactivity ratios. Using the value of r_2 as a "known" property of a copolymer, the decision tree attempted to predict r_1 since these predictions provided a lower error as opposed to disregarding r_2 (Figure 4-11). The justification for this was based on the relationships between the reactivity ratios that can be inferred from both the Mayo-Lewis model and the Q-e scheme. The copolymer mole fraction equation (equation 2) highlights that the feed of one monomer directly influences the reactivity of the other monomer and vice-versa.

Furthermore, the Q-e scheme displays similar behaviour in that the residual charge of the reacting groups of both monomers are intertwined during the determination of the reactivity ratios. Figure 4-11 supports the idea that the reactivity ratios are somehow related. Specifically, Mayo suggests that the dielectric constant of a radical is likely to play a pivotal role in radical polymerisation. Further improvements in future studies on applying machine learning to predict the reactivity ratios would

benefit from a more significant portion of dielectric constant values. Potentially, this would highlight the significance of the dielectric constant, indicating the importance of this intrinsic property.

Given the flaws in the total dataset used to train and predict the labels with the decision tree, interesting comparisons were observed with the Q-e scheme. First, there is already an uncertainty about the correct values of the r_1 and r_2 values. While carefully selected, the reality of creating a large database is that not every experiment is done by the same people or in the same circumstances. It is good to take the final predictions with a grain of salt and consider the value as an approximation. Second, the random forest technique relies on complete datasets. When establishing parameters for the monomers, one quickly discovers that, for example, not every monomer has a known (measured) dipole moment. For this reason, a lot of the data was generated with (relatively simple) Chemdraw software. These estimated dipole moments, refractive index, etc. are most likely not correct for the more complex molecules.

Using the predictions that considered r_2 and no minimum samples per leaf, the performance of the decision tree was placed against the Q-e scheme (Figure 4-4). The attempt of each model to predict r_1 consider different properties. As mentioned earlier, the Q-e utilises residual charges of radicals and general reactivities. Conversely, the decision tree solely considered the intrinsic properties of the monomers. Therefore, it is unsurprising that the Q-e scheme, an established, thoroughly researched and highly critiqued model, would generally perform better than the decision tree (Figure 4-6). Unfortunately, outperforming the Q-e scheme was not always the case. The anomalies at copolymer ID's 158 and 165 show a large percentage error. However, since the actual values of r_1 are small for those copolymers, minor deviations in the prediction lead to larger proportional errors. Notably, the predictions at ID's 175 and 178 showed improved predictions for r_1, and ID 152 had effectively the same percentage error for both predictions. This is a remarkable result. With large portions of data missing, a decision tree network could compete with the established Q-e scheme on the same predictions. Sourcing information on other values of Q and e proved a challenge. Future studies should investigate more values and continue these comparisons between future machine learning-based approaches at predictions and currently accepted Q-e model predictions.

4. Results and discussion

Similar to the findings of Chapter 3, the most challenging part of this study was finding sufficient data entries. An estimation of r_1 and r_2 values was made, but insufficient to show a significant improvement to the original Mayo-Lewis model and the Q-e scheme. Comparing to literature, there have been direct methods of calculating the reactivity rations. Johann et al. determined the activation energy via Density Function Theory (DTF) calculations, and then estimated the k_p values based on those outcomes.[110] In their work, they verify their calculations with real time NMR kinetics. Afterwards, they compared different experimental r_1 and r_2 values to the one obtained by their DFT calculations. The obtained results of r_1 and r_2 still show some deviation when compared to their experimental counterpart. Comparatively, the strategy attempted in this work was to not only obtain similar or better results, but also to avoid more complex calculations to reduce the computational cost of the calculations.

Webb et al. combined Coarse Grained (CG) modelling, ML, and model optimization to obtain a targeted sequence design.[83] Using coarse grained, molecular dynamic simulations, they could predict the structural properties of the polymer. In essence, they go directly from polymer sequence to predicting its properties. The method subscribed in this thesis makes use of physical parameters of the monomers in the polymer itself. The method from Webb et al. uses 2000 unique coarse-grained polymers. Important to note is that these data entries are all simulated by a Deep Neural Network (DNN). Again, this work was targeted to reduce the more complex calculations. Then again, the tested parameters were insufficient to get accurate predictions and further research is necessary. One of the potential stratifies is to try to use other descriptors such as the Norm-Index (*NI*) descriptor from Shi et al.[87], One-Hot Encoding (OHE) of the monomers, or even the parameters used in chapter 3. On top of that, it is worthwhile to revisit the strategy of this chapter as well, since the decision tree model might be too advanced (leading to overfitting) for the amount of datapoints.

In a recent study, published after the experiments from Chapter 4 were executed, Nguyen et al. successfully predicted the reactivity ratios using a ML model, Graph Attention Model, to predict the reactivity ratios based on the monomers' chemical structure.[111] Then again, it would be useful to gain knowledge about the effects of certain physical parameters of monomers on the copolymerisation.

5. Conclusion

Contemporary studies in the field of machine learning are advancing rapidly. Companies rely on fast, accurate, and reliable information to create new technologies and improve our lives. From autonomous vehicles, medicine, manufacturing, and even sports analytics, the scope of machine learning is vast. There have been limited studies that have attempted to combine polymer chemistry with machine learning. Commonly, they are used to predict desired properties of a polymer, such as high thermal conductivity for telecommunication. However, this paradigm is shifting. Greater urgency is being placed within the synthetic chemistry community to embrace the possibilities of machine learning in the field. This study has provided the basic framework to outline that, indeed, machine learning does have a vital role to play in the future of polymer chemistry. The predictions of the limited-data decision tree in this study were flawed. However, considering the difficulty in obtaining data, the predictions were not necessarily better or worse than an established model like the Q-e scheme. There was a clear trend that for copolymers with large values of r_1, the magnitude of error in the prediction of the decision tree was also relatively large. However, with the best value for the mean absolute error of only 5.46, there is a clear indication that the use of machine learning in some form can only improve, and perhaps one day, accurately predict the value of not only r_1 but r_2 as well.

Conclusion and Future Outlook

This project started out as research to find new, easy and reproducible methods to create dedicated polymer solubilizers but involved to something much more complex. In chapter 2, it was established that it is important to understand the relation between (a) a specific sequence and (b) the solubilizer capacities. When this relationship is understood, finding the proper sequence for a specific application will be easier, faster, and more efficient. However, determining the proper reaction conditions of a tailored polymer is an intensive and time-consuming process. An interesting new strategy to overcome this process is considering a synergy between classical polymer chemistry and computational science.

In order to bring the results of Chapter 2 closer to a so-called automatized research, the original approach needs to be restructured. One of the reasons batch reactions are often time consuming is that not only the reaction itself often takes longer, but sample preparation for every analysis technique is quite time consuming. As Van Herck et al. demonstrated, it is possible to do rapid kinetic screenings in continuous flow reactors.[112] Imagine a reactor setup where not only the screening is automated, but the purification and the analysis of their application (in this case drug solubility) was added in series as well. Thus, one could generate data in a fast and effective manner. Often when trying to optimise a certain synthesis, someone needs to make a calculated decision which parameters to vary to generate the desired outcome. The beautiful thing about machine learning is that this task can be automated as well. Sweidtmann et al. demonstrate how one can generate a Pareto front of multiple objectives. Often, when varying multiple input parameters, they are linked to each other (for example, increasing the temperature of a reaction will influence the speed of the reaction). A Pareto front displays the different sets of non-dominant solutions.[113] Optimising via such a Pareto front would simplify and accelerate the discovery of the optimal reactions. On top of this, generating a substantial amount of datapoints in a short time is an ideal base as input to use in machine learning. This becomes especially interesting if the database resulting from those experiments is not generated by one person, or even one lab, but by multiple scientists across the world. Large databases would need careful monitoring but getting consistent data is possible as demonstrated by a collaborative paper from Van Herck et al. where they proved that an operated independent experiment is possible.[114]

Where Chapter 2 focussed on “conventional” polymer chemistry, Chapter 3 and 4 focussed on the fundamentals of understanding and predicting the kinetic rate coefficient and reactivity ratios respectively. Logically, the more understanding there is about the polymerisations, the better “starting materials” an algorithm has to come to a final prediction. But, in order to fully understand the power of combining these two fields, one must first understand them separately. Hence, Chapter

3 and 4 are quite different from Chapter 2. There is hope that in future research, these two branches of science will be properly connected.

Overall, this thesis proves that there is a big value in combining knowledge of multiple fields to enhance the output. Logically, the correlation between monomers and their polymerisation descriptors exists but is just quite complex. Using machine learning, it is possible to estimate things that are normally too time consuming or too costly.

List of abbreviations

AIBN	Azobisisobutyronitrile
ATRP	Atom Transfer Radical Polymerization
BCS	Biopharmaceutics Classification System
BMA	*n*-Butyl Methacrylate
BP	British Pharmacopoeia
BPK	Boiling point (Kelvin)
CRP	Controlled Radical Polymerization
CTA	Chain Transfer Agent
DMAEMA	2-(Dimethylamino)ethyl methacrylate
DMF	Dimethylformamide
DP	Dipole moment
DSC	Differential Scanning Calorimetry
EPR	Electron paramagnetic resonance
FT-IR	Fourier-Transform Infrared spectroscopy
GAMESS	General Atomic and Molecular Electronic Structure System software package
GFE	Gibbs Free Energy
H	Hydrogen
IUPAC	International Union of Pure and Applied Chemistry
LASSO	Least Absolute Shrinkage and Selection Operator
LOOCV	Leave One Out Cross Validation
MePh	Toluene
MIPs	Molecularly Imprinted Polymers
ML	Machine Learning
MMA	Methyl methacrylate
MPC	2-Methacryloyloxyethyl Phosphorylcholine
MPK	Melting point (Kelvin)
MUMI	Multiple Unit Monomer Insertion
Mr	Molecular weight
NI	Norm Index
NIPAAM	poly(N-isopropylacrylamide-co-vinylpyrrolidone)
NME	New Molecular Entities
NMP	Nitroxide-Mediated Polymerization
PBS	Phosphate buffered saline
PCX	Piroxicam
PEG	Polyethylene glycol

PLP	Pulsed Laser Polymerization
QSPR	Quantitative Structure-Property Relationship
RAFT	Reversible Addition-Fragmentation chain-Transfer
RDRP	Reversible-Deactivation Radical Polymerization
SEC	Size Exclusion Chromatography
SUMI	Single Unit Monomer Insertion
THF	Tetrahydrofuran
USP	United States Pharmacopoeia
VP	1-Vinyl-2-pyrrolidone
XRD	X-Ray Diffraction
XRPD	X-Ray Powder Diffraction

Appendix A

1. Tables

Table A1: Alphabetic list of monomers considered with their abbreviation and SMILES notation.

Monomer name	Abbreviation	SMILES
1,3-Butadiene	**BuDE**	C=CC=C
2-(Hexylcarbamoyloxy)ethyl acrylate	**HCEA**	C=CC(OCCOC(NCCCCCC)=O)=O
2-(Hexylcarbamoyloxy)isopropyl acrylate	**HCPA**	C=C(OC(NCCCCCC)=O)C(OC(C)C)=O
2-(Phenylcarbamoyloxy)ethyl acrylate	**PhCEA**	C=CC(OCCOC(NC1=CC=CC=C1)=O)=O
2-(Phenylcarbamoyloxy)isopropyl acrylate	**PhCPA**	C=C(OC(NC1=CC=CC=C1)=O)C(OC(C)C)=O
2-ethylhexyl acrylate	**EHA**	C=CC(OC[C@H](CC)CCCC)=O
2-ethylhexyl methacrylate	**EHMA**	CC(C(OC[C@H](CC)CCCC)=O)=C
2-hydroxypropyl methacrylate	**HPMA**	CC(C(OC[C@H](O)C)=O)=C
2-propylheptyl acrylate	**PHA**	C=CC(OC[C@H](CCC)CCCCC)=O
Acrylonitrile	**CAN**	C=CC#N
Behenyl acrylate	**BeA**	C=CC(OCCCCCCCCCCCCCCCCCCCCCC)=O
Behenyl methacrylate	**BeMA**	CC(C(OCCCCCCCCCCCCCCCCCCCCCC)=O)=C
Benzyl acrylate	**BnA**	C=CC(OCC1=CC=CC=C1)=O
Benzyl methacrylate	**BzMA**	CC(C(OCC1=CC=CC=C1)=O)=C
Butyl methacrylate	**BMA**	CC(C(OCCCC)=O)=C
Cyclohexyl methacrylate	**CHMA**	CC(C(OC1CCCCC1)=O)=C
Dodecyl methacrylate	**DMA**	CC(C(OCCCCCCCCCCCC)=O)=C
Ethoxyethyl acrylate	**EEA**	C=CC(OCCOCC)=O
Ethyl methacrylate	**EMA**	CC(C(OCC)=O)=C
Glycidyl methacrylate	**GMA**	CC(C(OC[C@@H]1CO1)=O)=C
Henicosyl acrylate	**C21A**	C=CC(OCCCCCCCCCCCCCCCCCCCCC)=O
Heptadecyl acrylate	**C17A**	C=CC(OCCCCCCCCCCCCCCCCC)=O
Hydroxyethyl methacrylate	**HEMA**	CC(C(OCCO)=O)=C
Isobornyl acrylate	**iBoa**	C=CC(O[C@H]1C[C@@H]2CC[C@@]1(C)C2(C)C)=O
iso-bornyl methacrylate	**iBoMA**	CC(C(O[C@H]1C[C@@H]2CC[C@@]1(C)C2(C)C)=O)=C
iso-butyl methacrylate	**iBMA**	CC(C(OCC(C)C)=O)=C
iso-decyl methacrylate	**iDeMA**	CC(C(OCCCCCCCC(C)C)=O)=C
iso-nonyl acrylate	**INA-A**	C=CC(OCCCCCCC(C)C)=O
Methacrylic acid	**MAA**	CC(C(O)=O)=C
Methyl acrylate	**MA**	C=CC(OC)=O
Methyl methacrylate	**MMA**	CC(C(OC)=O)=C
n-Butyl acrylate	**BA**	C=CC(OCCCC)=O
n-Pentyl Methacrylate	**PnMA**	CC(C(OCCCCC)=O)=C
N-vinyl formamide	**NVF**	O=CNC=C
N-Vinyl Pyrrolidone	**NVP**	O=C1N(C=C)CCC1
Propylheptyl methacrylate	**PHMA**	CC(C(O[C@@H](CCC)CCCCCC)=O)=C
Stearyl acrylate	**SA**	C=CC(OCCCCCCCCCCCCCCCCCC)=O
Stearyl methacrylate	**SMA**	CC(C(OCCCCCCCCCCCCCCCCCC)=O)=C
Styrene	**Sty**	C=CC1=CC=CC=C1
tert-butyl acrylate	**tBA**	C=CC(OC(C)(C)C)=O
Vinyl acetate	**VAc**	CC(OC=C)=O

Table A2: List of monomers classified by type (acrylates, H-bonding monomers, methacrylates and other) with the activation energy E_a and pre-exponential factor A. Using the Arrhenius equation, the natural logarithm of the rate constant $\ln(k_p)$ was calculated for four different temperatures.

	Monomer	E_a / J·mol^{-1}	A / s^{-1}	$\ln(k_p$ / L·mol^{-1}·s^{-1}) T = 25°C	$\ln(k_p$ / L·mol^{-1}·s^{-1}) T = 50°C	$\ln(k_p$ / L·mol^{-1}·s^{-1}) T = 75°C	$\ln(k_p$ / L·mol^{1}·s^{-1}) T = 100°C	Source
Acrylates	**BA**	17.90	22100000	9.69	10.25	10.73	11.14	[115, 116]
	BeA	13.02	5350000	10.24	10.65	10.99	11.30	[116]
	BnA	16.12	12800000	9.86	10.36	10.80	11.17	[75]
	C17A	14.66	8150000	10.00	10.46	10.85	11.19	[116]
	C21A	12.99	3220000	9.74	10.15	10.50	10.80	[116]
	EEA	13.80	6300000	10.09	10.52	10.89	11.21	[116]
	EHA	15.80	9100000	9.65	10.14	10.57	10.93	[116]
	iBoa	15.35	4810000	9.19	9.67	10.08	10.44	[75]
	INA-A	16.54	13500000	9.75	10.26	10.70	11.09	[116]
	MA	17.30	14100000	9.48	10.02	10.48	10.89	[117]
	PHA	16.41	10500000	9.55	10.06	10.50	10.88	[116]
	SA	16.93	18600000	9.91	10.44	10.89	11.28	[116]
	***t*BA**	18.90	22100000	9.69	10.25	10.73	11.14	[118]
H-Bonding monomers	**HCEA**	13.30	6600000	10.34	10.75	11.11	11.42	[119]
	HCPA	14.10	6600000	10.01	10.45	10.83	11.16	[119]
	HEMA	21.90	8880000	7.16	7.85	8.43	8.94	[120]
	PhCEA	14.30	12000000	10.53	10.98	11.36	11.69	[119]
	PhCPA	14.20	4900000	9.68	10.12	10.50	10.83	[119]
Methacrylates	**BeMA**	20.52	2510000	6.46	7.10	7.65	8.12	[72]
	BMA	22.90	3801894	5.91	6.63	7.24	7.77	[121]
	BzMA	22.90	6760830	6.49	7.20	7.82	8.35	[122]
	CHMA	23.00	6309573	6.38	7.10	7.71	8.24	[122]
	DMA	21.00	2511886	6.27	6.92	7.48	7.97	[121]
	EHMA	21.60	2390000	5.97	6.65	7.22	7.72	[120]
	EMA	23.40	4073803	5.78	6.51	7.14	7.68	[121]
	GMA	22.90	5011872	6.19	6.90	7.52	8.05	[122]
	HPMA	20.80	3510000	6.68	7.33	7.88	8.37	[120]
	iBMA	21.80	2640000	5.99	6.67	7.25	7.76	[120]
	iBoMA	23.10	6165950	6.32	7.04	7.65	8.19	[122]
	iDeMA	21.60	2390000	5.97	6.65	7.22	7.72	[120]
	MMA	22.36	2673006	5.78	6.48	7.07	7.59	[123]
	PHMA	21.72	2830000	6.09	6.77	7.35	7.85	[72]
	PnMA	23.80	6000000	6.01	6.75	7.38	7.94	[124]
	SMA	21.49	3450000	6.39	7.06	7.63	8.13	[72]
Other	**BuDE**	35.70	80500000	3.81	4.91	5.87	6.70	[120]
	CAN	15.40	1790000	8.19	8.67	9.08	9.43	[125]
	MAA	16.10	380000	6.35	6.86	7.29	7.66	[120]
	NVF	19.50	6400000	7.81	8.41	8.93	9.39	[126]
	NVP	17.60	25700000	9.96	10.51	10.98	11.39	[127]
	Sty	32.51	42657952	4.45	5.47	6.34	7.09	[128]
	VAc	20.40	13500000	8.19	8.83	9.37	9.84	[129]

Table A3: List with monomer parameters classified by type (acrylates, H-bonding monomers, methacrylates and other). The molecular weight *MW*, dipole moment *DP*, boiling point *BP* and the Gibbs free energy *GFE* were predicted with chemdraw. The other parameters were researched from literature. The A_value is the substituent effect on the C2 carbon of the polymerising double bond, the R_value is the substituent effect on the moiety chain.

	Monomer	MW / g·mol^{-1}	DP	BP / K	GFE / J	A_value [130]	R_value [130]
Acrylates	**BA**	128.17	2.27	397	-273.05	0.00	11.28
	BeA	380.60	2.28	691	-124.33	0.00	12.67
	BnA	162.18	2.40	506	-138.22	0.00	13.08
	C17A	310.50	3.17	633	-166.43	0.00	12.67
	C21A	366.60	2.76	679	-132.75	0.00	12.71
	EEA	144.17	3.91	442	-380.89	0.00	13.43
	EHA	184.27	2.96	490	-244.65	0.00	12.46
	iBoA	208.30	2.37	516	-142.37	0.00	12.42
	INA-A	198.30	2.50	509	-236.23	0.00	12.71
	MA	86.09	2.23	349	-301.15	0.00	12.71
	PHA	212.33	2.50	527	-227.81	0.00	12.44
	SA	324.50	2.67	644	-158.01	0.00	12.71
	***t*BA**	128.17	2.27	397	-273.05	0.00	11.28
H-Bonding monomers	**HCEA**	243.30	1.74	585	-454.77	0.00	13.79
	HCPA	257.33	4.94	587	-457.34	1.91	12.15
	HEMA	130.14	3.68	472	-429.68	-1.55	13.23
	HPMA	144.17	3.43	473	-423.70	-1.55	13.00
	PhCEA	235.24	1.07	602	-342.36	0.00	13.70
	PhCPA	249.26	4.29	603	-344.93	2.51	12.15
Methacrylates	**BeMA**	394.70	1.65	699	-124.46	-1.55	12.67
	BMA	142.20	1.71	436	-276.02	-1.55	13.24
	BzMA	176.21	1.71	519	-138.35	-1.55	13.08
	CHMA	168.23	1.27	486	-234.73	-1.55	12.29
	DMA	254.41	1.50	579	-208.66	-1.55	12.71
	EHMA	198.30	1.59	504	-244.78	-1.55	12.46
	EMA	114.12	1.70	391	-292.86	-1.55	12.58
	GMA	142.15	4.10	445	-309.81	-1.55	13.12
	iBMA	142.20	1.72	423	-278.46	-1.55	12.50
	iBoMA	222.32	1.85	529	-142.50	-1.55	12.42
	iDeMA	226.35	1.61	539	-227.94	-1.55	12.67
	MMA	100.12	1.64	367	-301.28	-1.55	12.71
	PHMA	226.35	1.51	539	-227.94	-1.55	12.09
	PnMA	156.2	1.65	457	-267.60	-1.55	12.75
	SMA	338.6	1.96	653	-158.14	-1.55	12.71
Other	**BuDE**	54.09	0.00	289	97.06	0.00	0.00
	CAN	53.06	3.94	368	133.98	0.00	14.33
	MAA	86.09	1.78	433	-343.24	-1.55	12.36
	NVF	71.08	3.98	377	-9.33	-1.55	3.48
	NVP	111.14	3.95	493	70.61	0.00	-0.33
	Sty	104.15	0.26	420	155.31	0.00	1.76
	VAc	86.09	1.54	349	-301.15	0.00	3.90

Table A4: List with monomer parameters classified by type (acrylates, H-bonding monomers, methacrylates and other). H_acc and H_don are the H-bonding effects, A1* and A2* are the dissociation constants. The values of these parameters were taken from literature.

	Monomer	H_acc [131]	H_don [131]	A1* [132]	A2*
Acrylates	**BA**	0.0	-0.5	-2.8	1
	BeA	0.0	-0.5	-2.8	1
	BnA	-2.6	0.0	-2.8	1
	C17A	0.0	-0.5	-2.8	1
	C21A	0.0	-0.5	-2.8	1
	EEA	-5.1	0.0	-2.8	1
	EHA	0.0	-0.5	-2.8	1
	iBoA	0.0	-0.5	-2.8	1
	INA-A	0.0	-0.5	-2.8	1
	MA	0.0	-0.5	-2.8	1
	PHA	0.0	-0.5	-2.8	1
	SA	0.0	-0.5	-2.8	1
	***t*BA**	0.0	-0.5	-2.8	1
H-Bonding monomers	**HCEA**	-7.7	-4.1	-2.8	1
	HCPA	-16.8	-4.1	-2.8	1
	HEMA	-5.5	-3.5	-2.8	647
	HPMA	-5.5	-3.5	-2.8	647
	PhCEA	-11.0	-3.7	-2.8	1
	PhCPA	-13.1	-3.7	-2.8	1
Methacrylates	**BeMA**	0.0	-0.5	-2.8	647
	BMA	0.0	-0.5	-2.8	647
	BzMA	-2.6	0.0	-2.8	647
	CHMA	0.0	-0.5	-2.8	647
	DMA	0.0	-0.5	-2.8	647
	EHMA	0.0	-0.5	-2.8	647
	EMA	0.0	-0.5	-2.8	647
	GMA	-5.1	0.0	-2.8	647
	iBMA	0.0	-0.5	-2.8	647
	iBoMA	0.0	-0.5	-2.8	647
	iDeMA	0.0	-0.5	-2.8	647
	MMA	0.0	-0.5	-2.8	647
	PHMA	0.0	-0.5	-2.8	647
	PnMA	0.0	-0.5	-2.8	647
	SMA	0.0	-0.5	-2.8	647
Other	**BuDE**	0.0	0.0	0.0	-
	CAN	-4.7	-4.7	-3.4	10
	MAA	0.0	-5.0	-2.8	-
	NVF	-6.9	-1.6	-3.4	-
	NVP	-5.5	-4.1	-3.4	-
	Sty	-2.6	0.0	-8.4	15
	VAc	-2.9	0.0	-5.9	-

Table A5: List with monomer parameters from Chemspider, generated using the ACD/Labs Percepta Platform – PhysChem Module. The selected parameters are the boiling point *BP*, the refraction index *RI*, the density *ρ,* the partition coefficient ACDlogP, the polarisation, the surface tension and the vapor pressure.

	Monomer	BP / ˚C	RI	ρ / g·mL^{-1}	ACDlogP	Polarisation	Surface tension	Vapor pressure
Acrylates	**BA**	145.0	1.418	0.9	2.39	14.3	26.7	4.8
	BeA	455.7	1.455	0.9	11.95	47.3	31.3	0.0
	BnA	228.7	1.517	1.1	2.27	18.5	36.1	0.1
	C17A	385.6	1.451	0.9	9.30	38.1	30.8	0.0
	C21A	442.3	1.455	0.9	11.42	45.5	31.2	0.0
	EEA	174.6	1.420	1.0	1.02	15.0	28.4	1.2
	EHA	216.0	1.434	0.9	4.33	21.6	28.0	0.1
	iBoA	244.5	1.491	1.0	4.22	23.7	33.0	0.0
	INA-A	247.3	1.437	0.9	4.86	23.4	28.4	0.0
	MA	80.2	1.390	0.9	0.79	8.8	23.5	86.3
	PHA	266.1	1.439	0.9	5.39	25.3	28.7	0.0
	SA	400.2	1.452	0.9	9.83	40.0	30.9	0.0
	***t*BA**	133.0	1.418	0.9	2.02	14.3	25.1	8.6
H-Bonding monomers	**HCEA**	357.3	1.456	1.0	3.43	25.6	34.6	0.0
	HCPA	366.2	1.456	1.0	3.78	27.4	33.8	0.0
	HEMA	189.0	1.443	1.1	0.50	13.0	33.7	0.2
	HPMA	189.0	1.458	1.1	1.39	14.0	36.1	0.6
	PhCEA	218.8	1.444	1.0	0.85	14.8	32.3	0.0
	PhCPA	322.2	1.557	1.2	2.65	24.7	45.9	0.0
Methacrylates	**BeMA**	331.7	1.549	1.2	3.00	26.5	43.7	0.0
	BMA	468.7	1.455	0.9	12.5	49.1	31.0	0.0
	BzMA	160.0	1.423	0.9	2.94	16.0	26.4	2.4
	CHMA	247.0	1.512	1.0	2.82	20.2	34.8	1.0
	DMA	210.0	1.459	1.0	3.40	18.9	30.5	0.2
	EHMA	322.7	1.445	0.9	7.19	30.7	29.6	0.0
	EMA	234.8	1.436	0.9	4.88	23.4	27.7	0.1
	GMA	120.5	1.410	0.9	1.88	12.4	24.7	15.2
	iBMA	155.0	1.421	0.9	2.76	16.0	25.4	3.1
	iBoMA	263.1	1.488	1.0	4.77	25.4	32.4	0.0
	iDeMA	283.9	1.441	0.9	5.94	27.0	28.4	0.0
	MMA	100.3	1.400	0.9	1.35	10.5	23.5	36.9
	PHMA	-	-	-	-	-	-	-
	PnMA	191.3	1.427	0.9	3.47	17.9	27.0	0.5
	SMA	414.3	1.452	0.9	10.38	41.8	30.6	0.0
Other	**BuDE**	-4.4	1.389	0.6	1.86	7.9	15.9	2101
	CAN	77.3	1.385	0.8	0.19	6.2	25.0	97.1
	MAA	160.5	1.430	1.0	0.83	8.7	30.9	1.2
	NVF	184.3	1.406	0.9	-0.18	7.7	25.0	0.7
	NVP	217.6	1.593	1.1	0.37	13.1	52.4	0.1
	Sty	145.2	1.558	0.9	2.70	14.7	31.0	6.2
	Vac	72.5	1.390	0.9	0.73	8.8	23.5	118.5

Table A6: Results of Figure 3-6c

	Monomer	Experimental $\ln(k_p$ / L·mol^{-1}·s$^{-1})$ T = 25˚C	Predicted $\ln(k_p$ / L·mol^{-1}·s$^{-1})$ T = 25˚C	Relative error [experimental – predicted] $\ln(k_p$ / L·mol^{-1}·s$^{-1})$ T = 25˚C
Acrylates	**BA**	9.689923	9.458881	0.231
	SA	9.908823	9.842755	0.066
	PHA	9.546813	9.659156	-0.110
	MA	9.482579	9.381649	0.101
	INA-A	9.745663	9.606882	0.139
	iBoA	9.193703	9.757866	-0.560
	***t*BA**	9.689923	9.630068	0.060
	EEA	10.088888	9.873488	0.215
	C21A	9.744492	9.925102	-0.180
	C17A	9.999434	9.827086	0.172
	BnA	9.861884	9.677131	0.185
	BeA	10.240103	9.960121	0.280
	EHA	9.649756	9.605180	0.045
H-Bonding monomers	**PhCPA**	9.676210	9.593760	0.082
	PhCEA	10.531536	10.621373	-0.090
	HPMA	6.679599	6.503095	0.177
	HCPA	10.014403	10.193138	-0.180
	HCEA	10.337119	10.430283	-0.090
	HEMA	7.164720	6.324925	0.840
Methacrylates	**iBoMA**	6.315358	6.401280	-0.090
	SMA	6.385194	6.046577	0.339
	PnMA	6.006353	5.948854	0.057
	MMA	5.777652	6.199693	-0.420
	iDeMA	5.973810	6.055019	-0.080
	iBMA	5.991465	5.969650	0.022
	BMA	5.913503	5.983812	-0.070
	EMA	5.780744	6.162680	-0.380
	EHMA	5.973810	6.246495	-0.270
	DMA	6.265301	6.178625	0.087
	CHMA	6.378426	6.244365	0.134
	BzMA	6.487684	6.518645	-0.030
	BeMA	6.458338	6.504005	-0.050
	GMA	6.188264	6.324925	-0.140
Other	**Sty**	4.454347	4.599294	-0.140
	CAN	8.185071	8.590136	-0.410

Table A7: List with monomer collected experimental parameters from Chemspider, including the boiling point *BP*, refraction index *RI* and the density *ρ*. The A2* dissociation constant parameter is included as well. Monomers that had a value for all of four categories were highlighted in bold and coloured according to their coloration used Figure 4-23.

	Monomer	BP / ˚C	RI	ρ / g·mL^{-1}	A2*
Acrylates	**BA**	**145**	**1.418**	**0.895**	**1**
	BeA	-	-	0	1
	BnA	111	-	1.060	1
	EEA	174	-	-	1
	iBoa	119	1.476	-	1
	MA	**80.5**	**1.403**	**0.956**	**1**
	***t*BA**	**121**	**1.411**	**0.883**	**1**
H-bonding monomers	**HEMA**	**250**	**1.453**	**1.071**	**647**
Methacrylates	**BzMA**	**162**	**1.424**	**0.894**	**647**
	CHMA	**231**	**1.514**	**1.040**	**647**
	DMA	**72**	**1.457**	**0.970**	**647**
	EHMA	**142**	**1.445**	**0.873**	**647**
	GMA	**118**	**1.413**	**0.917**	**647**
	iBMA	**155**	**1.420**	**0.889**	**647**
	MMA	**100**	**1.414**	**0.939**	**647**
	SMA	-	-	0.860	647
Other	**BuDE**	-4.5	1.429	0.620	-
	CAN	**77**	**1.391**	**0.806**	**10**
	MAA	163	1.431	1.015	-
	NVF	210	1.494	1.014	-
	NVP	92	1.512	1.044	-
	Sty	**145**	**1.546**	**0.906**	**15**
	VAc	72	1.395	0.932	-

Table A8: Coefficients of the model with acrylates and methacrylates

Parameters	Standard Deviation	Coefficients	Importance
Polarizability	808.4313	0.00268	2.166596
Dissociation constant 2	328.9271	-0.00555	1.825545
Molecular weight	89.20704	0.00149	0.132918
Effect of H-donor	1.477476	-0.06087	0.089934
Inductive effect of tailgroup	0.384931	0.00162	0.000624
Effect of H-acceptor	0.181007	0.01008	0.001825
Inductive effect	0.790796	0.00001	7.91E-06
Dissociation constant 1	9.05E-16	0	0

Table A9: Coefficients of the model with acrylates, methacrylates and H-bonding monomers

Parameters	Standard Deviation	Coefficients	Importance
Dissociation constant 2	327.8574	-0.00636	2.085173
Polarizability	10.96917	0.10146	1.112932
Molecular weight	83.51009	-0.01179	0.984584
Effect of H-acceptor	4.292892	-0.10600	0.455047
Inductive effect	1.058983	-0.37847	0.400793
Effect of H-donor	1.323033	-0.07107	0.094028
Inductive effect of tailgroup	0.468596	0.08725	0.040885
Dissociation constant 1	1.35E-15	0	0

Table A10: Coefficients of the model with acrylates, methacrylates and others

Parameters	Standard Deviation	Coefficients	Importance
Dissociation constant 2	327.7552	-0.00526	1.723992
Molecular weight	92.44408	0.01385	1.280351
Polarizability	11.97967	-0.08162	0.977781
Inductive effect of tailgroup	2.09159	0.17573	0.367555
Dissociation constant 1	1.041882	0.10051	0.104720
Effect of H-acceptor	1.65423	0.02534	0.041918
Effect of H-donor	0.819197	0.02161	0.017703
Inductive effect	0.789821	-0.00071	0.000561

Table A11: Coefficients of the model with acrylates, methacrylates, H-bonding monomers and others

Parameters	Standard Deviation	Coefficients	Importance
Dissociation constant 2	325.9089	-0.00676	2.203144
Molecular weight	86.79396	0.01508	1.308853
Polarizability	11.15569	-0.10046	1.120701
Dissociation constant 1	0.94902	1.09465	1.038845
Inductive effect	1.037659	-0.57508	0.596737
Inductive effect of tailgroup	1.934587	-0.13211	0.255578
Effect of H-donor	1.438212	-0.06727	0.096749
Effect of H-acceptor	4.185084	-0.00045	0.001883

2. Code

Import the libraries

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

from sklearn.linear_model import LinearRegression, RidgeCV, LassoCV, Ridge,
Lasso
from sklearn.model_selection import cross_validate, cross_val_predict,
cross_val_score, LeaveOneOut, train_test_split, GridSearchCV
from sklearn.metrics import r2_score, confusion_matrix, mean_squared_error
from sklearn.base import BaseEstimator
```

Definitions

```
# Calculation r2
def r2_from_sqerrs_and_values(squared_errors, y):
    return 1-squared_errors.mean()/(y.var())

# Calculation Mean Squared Error
def MSE_from_sqerrs_and_values(squared_errors, y):
    return -squared_errors.mean()

# Calculation Error
def error_from_sqerrs_and_values(squared_errors, y):
    return (np.exp(np.sqrt(-squared_errors.mean()))-1)*100

# Calculation Squared Error
def squared_errors(squared_errors, y):
    return np.sqrt(-1*squared_errors)


def metrics(predicted, labels):
    RMSE = mean_squared_error(labels,predicted, squared=False)
    R2 = r2_score(labels,predicted)
    var_pred = predicted.var()
    var_res = (labels-predicted).var()
    print("Root mean Squared Error is", RMSE)
    print("The R2 value is", R2)
    print("Variance of predictions is", var_pred)
    print("Variance of residuals is", var_res)
```

Import the “Kp Database.csv” file and sort the database by type

```
# Import the database csv file
df = pd.read_csv('Kp Database.csv', encoding = 'ISO-8859-1')

# Do a sorting of the database and reset the index number
df = df.sort_values(by='Type')

# Print the first few entries of the dataset
print(df.head(5))
print(df)
```

Create a data frame

```
# OPTION 1: Create data frame and select the parameters (columns)

data =
df[['Weight_zero','Weight_equal','Weight_half','Spider_BP_P','Spider_RI_P','S
pider_D_P','Spider_ACDlogP','Spider_P','Spider_ST','Spider_VP',
'Monomer','A','Acrylate','Ea', 'Type','R', 'Mr', 'DP', 'BPK', 'MPK',
'GFE','Kp25','Kp50','Kp75','Kp100', 'ln(Kp25)', 'ln(Kp50)', 'ln(Kp75)',
'ln(Kp100)', 'a_value', 'Hbond', 'Hacc', 'Hdon', 'A1*', 'A2*']]

# OPTION 2: This selection has more parameters, but also more empty cells (n
= 13)
data =
df[['Weight_equal','Spider_BP','Spider_BP_P','Spider_RI','Spider_RI_P','Spide
r_D','Spider_D_P','Spider_ACDlogP','Spider_P','Spider_ST','Spider_VP',
'Monomer','A','Acrylate','Ea', 'Type','R', 'Mr', 'DP', 'BPK', 'MPK',
'GFE','Kp25','Kp50','Kp75','Kp100', 'ln(Kp25)', 'ln(Kp50)', 'ln(Kp75)',
'ln(Kp100)', 'a_value', 'Hbond', 'Hacc', 'Hdon', 'A1*', 'A2*']]


# Further selection of the data by type

# OPTION 1.1/2.1
data = data[(data.Type == 'Acrylate') | (data.Type == 'Methacrylate')|
(data.Type == 'H-bonding monomers')]
# OPTION 1.2/2.2
data = data[(data.Type == 'Acrylate') | (data.Type == 'Methacrylate')|
(data.Type == 'Other')]
# OPTION 1.3/2.3
data = data[(data.Type == 'Acrylate') | (data.Type == 'Methacrylate')]

# Leave out all the rows that do not have an entry
data = data.dropna(how = 'any',axis = 0).reset_index(drop = True)

# Reset the index
data = data.reset_index()
```

Select the X and Y

```
# Select the parameters for X; THESE PARAMETERS CAN EASILY BE VARIED, BELOW
IS ONE EXAMPLE FOR ILLUSTRATION

X_train = data[['Mr','a_value', 'R', 'Hacc', 'Hdon','A1*', 'A2*',
'Spider_P']]

#OPTION 1: Select the parameters for y = ln(Kp)
y_train = data['ln(Kp25)']
y_train25 = data['ln(Kp25)']
y_train50 = data['ln(Kp50)']
y_train75 = data['ln(Kp75)']
y_train100 = data['ln(Kp100)']

#Option 2: Select the parameters for y = Kp
y_train = data['Kp25']
y_train25 = data['Kp25']
y_train50 = data['Kp50']
y_train75 = data['Kp75']
y_train100 = data['Kp100']
```

Do a regression for every temperature, the **sample weight can be varied as well**

```
# Result for 25 degrees

res25 = cross_validate(RidgeCV(alphas=(10**np.linspace(10,-2,300)*0.5)),
X_train, y_train25, cv=LeaveOneOut(), scoring='neg_mean_squared_error',
return_train_score=True, return_estimator=True, fit_params={'sample_weight':
data.Weight_equal})


sum = np.zeros(len(res25["estimator"]))

for est in res25["estimator"]:
    sum = sum + est.predict(X_train)

rid_predsR25 = sum / len(res25["estimator"])
#print(rid_predsR25)

Result25 = pd.DataFrame()
Result25['25 degrees experimental'] = y_train25
Result25['25 degrees predicted'] = rid_predsR25
Result25['25 degrees type'] = data['Type']
print(Result25)
```

```
# Result for 50 degrees

res50 = cross_validate(RidgeCV(alphas=(10**np.linspace(10,-2,300)*0.5)),
X_train, y_train50, cv=LeaveOneOut(), scoring='neg_mean_squared_error',
return_train_score=True, return_estimator=True, fit_params={'sample_weight':
data.Weight_equal})

sum = np.zeros(len(res50["estimator"]))

for est in res50["estimator"]:
    sum = sum + est.predict(X_train)

rid_predsR50 = sum / len(res50["estimator"])
#print(rid_predsR50)

Result50 = pd.DataFrame()
Result50['50 degrees experimental'] = y_train50
Result50['50 degrees predicted'] = rid_predsR50
print(Result50)


# Result for 75 degrees
res75 = cross_validate(RidgeCV(alphas=(10**np.linspace(10,-2,300)*0.5)),
X_train, y_train75, cv=LeaveOneOut(), scoring='neg_mean_squared_error',
return_train_score=True, return_estimator=True, fit_params={'sample_weight':
data.Weight_equal})

sum = np.zeros(len(res75["estimator"]))

for est in res75["estimator"]:
    sum = sum + est.predict(X_train)

rid_predsR75 = sum / len(res75["estimator"])
#print(rid_predsR75)

Result75 = pd.DataFrame()
Result75['75 degrees experimental'] = y_train75
Result75['75 degrees predicted'] = rid_predsR75
print(Result75)
```

```
# Result for 100 degrees
res100 = cross_validate(RidgeCV(alphas=(10**np.linspace(10,-2,300)*0.5)),
X_train, y_train100, cv=LeaveOneOut(), scoring='neg_mean_squared_error',
return_train_score=True, return_estimator=True, fit_params={'sample_weight':
data.Weight_equal})

sum = np.zeros(len(res100["estimator"]))

for est in res100["estimator"]:
    sum = sum + est.predict(X_train)

rid_predsR100 = sum / len(res100["estimator"])
#print(rid_predsR100)

Result100 = pd.DataFrame()
Result100['100 degrees experimental'] = y_train100
Result100['100 degrees predicted'] = rid_predsR100
print(Result100)
```

Calculate the metrics

```
metrics(rid_predsR25, y_train25)
metrics(rid_predsR50, y_train50)
metrics(rid_predsR75, y_train75)
metrics(rid_predsR100, y_train100)
```

Make an Arrhenius plot and calculate the slope and activation energy

```
R_slope_list = []
R_intercept_list = []

rid_predsR25_list = rid_predsR25.tolist()
rid_predsR50_list = rid_predsR50.tolist()
rid_predsR75_list = rid_predsR75.tolist()
rid_predsR100_list = rid_predsR100.tolist()

for (i, j, k, l) in zip(rid_predsR25_list, rid_predsR50_list,
rid_predsR75_list, rid_predsR100_list):
    y = [i, j, k, l]
    x = [1/(25+273.15), 1/(50+273.15), 1/(75+273.15), 1/(100+273.15)]
    slope, intercept = np.polyfit(x,y,1)
    R_slope_list.append(slope)
    R_intercept_list.append(intercept)

df_R_s = pd.DataFrame({'Slope Predicted': R_slope_list})
df_R_s['Ea Experimental'] = data[['Ea']]
df_R_s['Ea Predicted'] = -df_R_s['Slope Predicted']*8.314/1000
df_R_s['ln(Ea Predicted)'] = np.log(-df_R_s['Slope Predicted']*8.314/1000)
df_R_s['Type'] = data[['Type']]
df_R_s['Name'] = data[['Monomer']]
df_R_s['Acrylate'] = data[['Acrylate']]

df_R_i = pd.DataFrame({'Intercept Predicted': R_intercept_list})
df_R_i['A Experimental'] = data[['A']]
df_R_i['A Predicted'] = np.exp(df_R_i['Intercept Predicted'])
df_R_i['ln(A Predicted)'] = np.log(df_R_i['A Predicted'])
df_R_i['ln(A Experimental)'] = np.log(df_R_i['A Experimental'])
df_R_i['Type'] = data[['Type']]
df_R_i['Acrylate'] = data[['Acrylate']]

print(df_R_s)
print(df_R_i)
print(df_R_i['ln(A Experimental)'])
```

Calculate the metrics

```
metrics(df_R_s['Ea Predicted'], df_R_s['Ea Experimental'])
metrics(df_R_i['ln(A Predicted)'], df_R_i['ln(A Experimental)'])
```

Import the train dataset

```
df = pd.read_csv('Kp Database.csv', encoding = 'ISO-8859-1')
df = df.sort_values(by='Type') #sort by type to make copypaste easier
df = df.reset_index()

data = df[['Type', 'Mr', 'DP', 'BPK', 'GFE', 'R', 'Hacc', 'Hdon', 'A1*',
'A2*', 'Spider_P', 'ln(Kp25)', 'a_value']]

data = data.dropna(how = 'any',axis = 0).reset_index(drop = True)
data = data.reset_index()
print(data)

X_train = data[['Mr', 'a_value','R', 'Hacc', 'Hdon', 'A1*', 'A2*',
'Spider_P']]
y_train = data['ln(Kp25)']
X_test = df_test[['Mr', 'a_value', 'R', 'Hacc', 'Hdon', 'A1*', 'A2*',
'Spider_P']]
```

Calculate the coefficients of the model

```
coefs = []
ridgeCV = RidgeCV(alphas=(10**np.linspace(10,-2,300)*0.5)).fit(X_train,
y_train)
ridgeCV.coef_
coefs.append(ridgeCV.coef_)
print(coefs)


# define the model
model = RidgeCV(alphas=(10**np.linspace(10,-2,300)*0.5)).fit(X_train,
y_train25)
# fit the model
model.fit(X_train, y_train25)
# get importance
importance = model.coef_
# summarize feature importance
for i,v in enumerate(importance):
 print('Feature: %0d, Score: %.5f' % (i,v))
# plot feature importance
pyplot.bar([x for x in range(len(importance))], importance)
pyplot.show()

coefs = []
coefs.append(ridgeCV.coef_)
print(coefs)
```

3. Plots

The resulting plots before and after cross validation are variations of the same principle. The distinction in each plot is made clear in the caption. Note the difference in:

- The selected data: n = amount of monomers in the chosen dataset
- Which parameter is predicted: $\ln(k_p)$ or k_p
- Which regression was chosen: Linear, Lasso or Ridge
- Which temperature was used
- What the parameters were
- Whether or not a weight was added to certain parts of the data

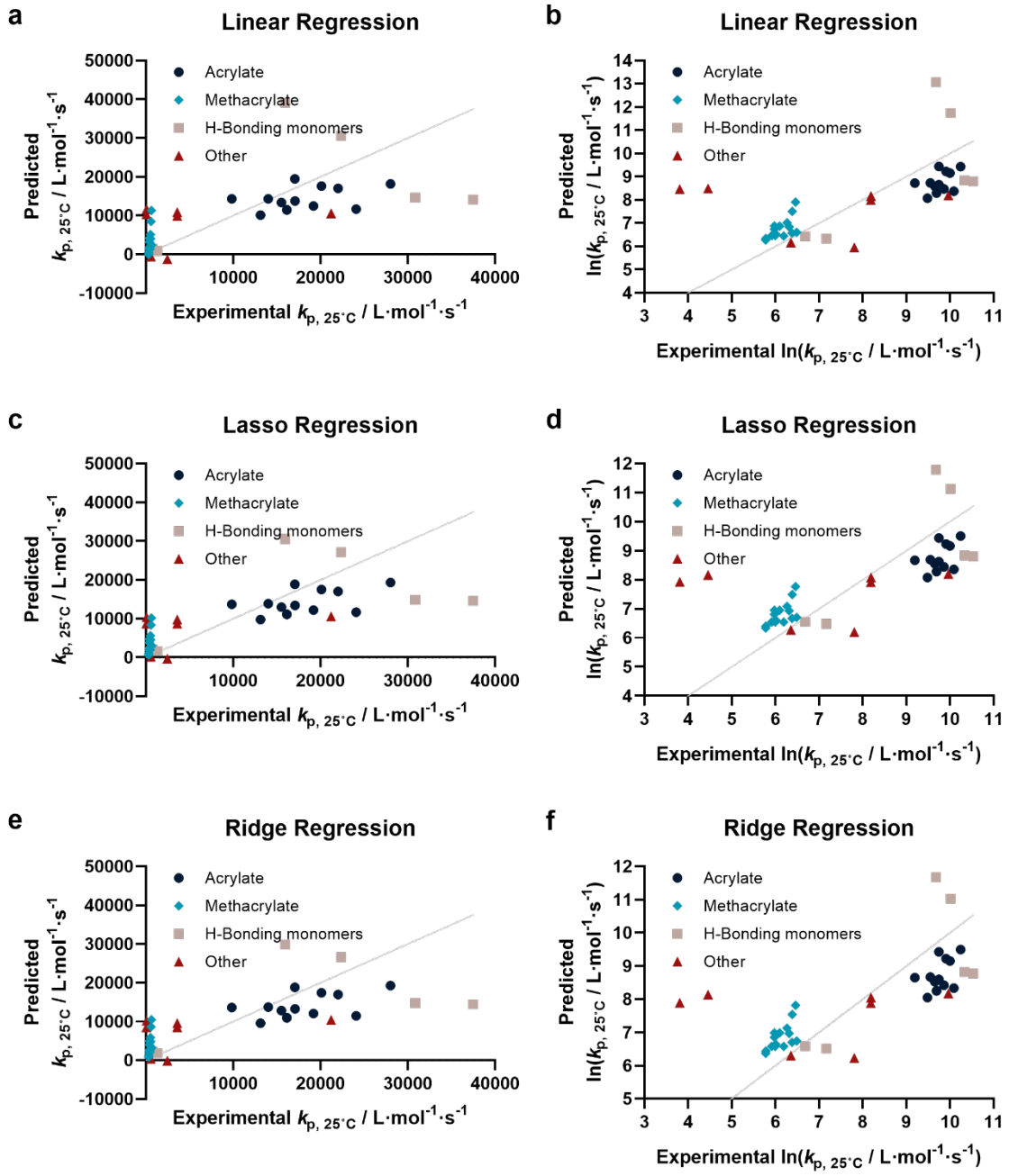


Figure 4-7: Correlation plots of the predicted propagation rate constant (k_p) and the natural logarithm of the predicted propagation rate constant (ln(k_p)) at 25˚C versus their experimental values for different regressions. Linear (a and b), Lasso (c and d) and Ridge (e and f) regressions were done using all available data (n = 41). The predictions were generated with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers) as parameters. The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	7.892E3	1.470	7.085E3	1.330	7.114E3	1.334
R2 Value	0.434	0.414	0.544	0.521	0.054	0.518
Variance Predictions	6.865E7	2.167	5.053E7	1.645	4.782E7	1.563
Variance Residuals	6.380E7	2.213	5.145E7	1.813	5.187E7	1.824

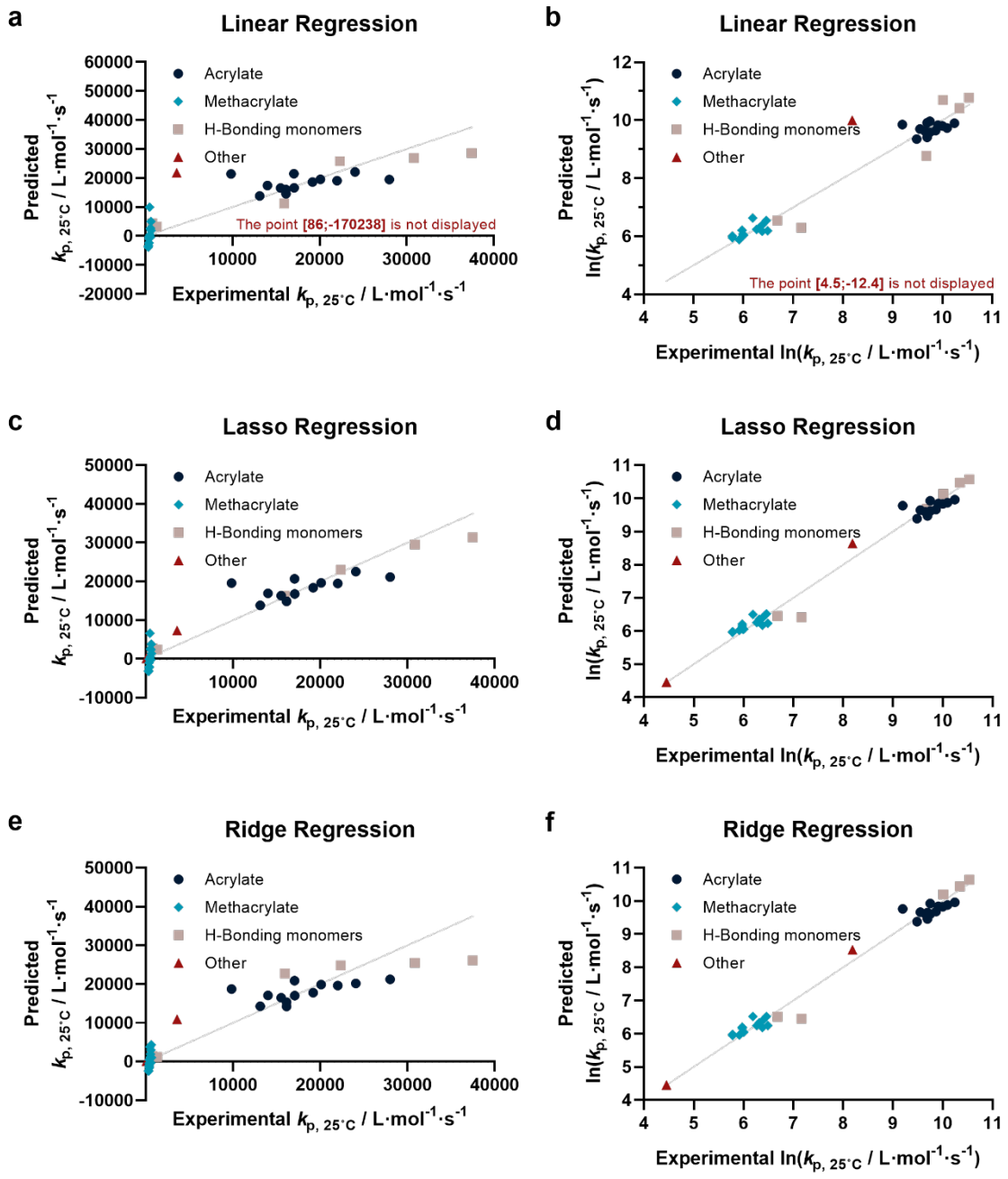


Figure 4-8: Correlation plots of the predicted propagation rate constant (k_p) and the natural logarithm of the predicted propagation rate constant (ln(k_p)) at 25˚C versus their experimental values for different regressions. The Linear (a and b), Lasso (c and d) and Ridge (e and f) regression were done using all available data (n = 35). The predictions were generated with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, the effect of H-donors and H-acceptors, and the dissociation constants as parameters. The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	2.924E4	2.889	3.113E3	0.235	3.784E3	0.223
R2 Value	-6.420	-1.350	0.916	0.984	0.876	0.986
Variance Predictions	1.010E9	15.005	1.039E8	3.471	9.580E7	3.484
Variance Residuals	8.595E8	8.386	9.977E6	0.057	1.474E7	0.051

Results after Cross Validation

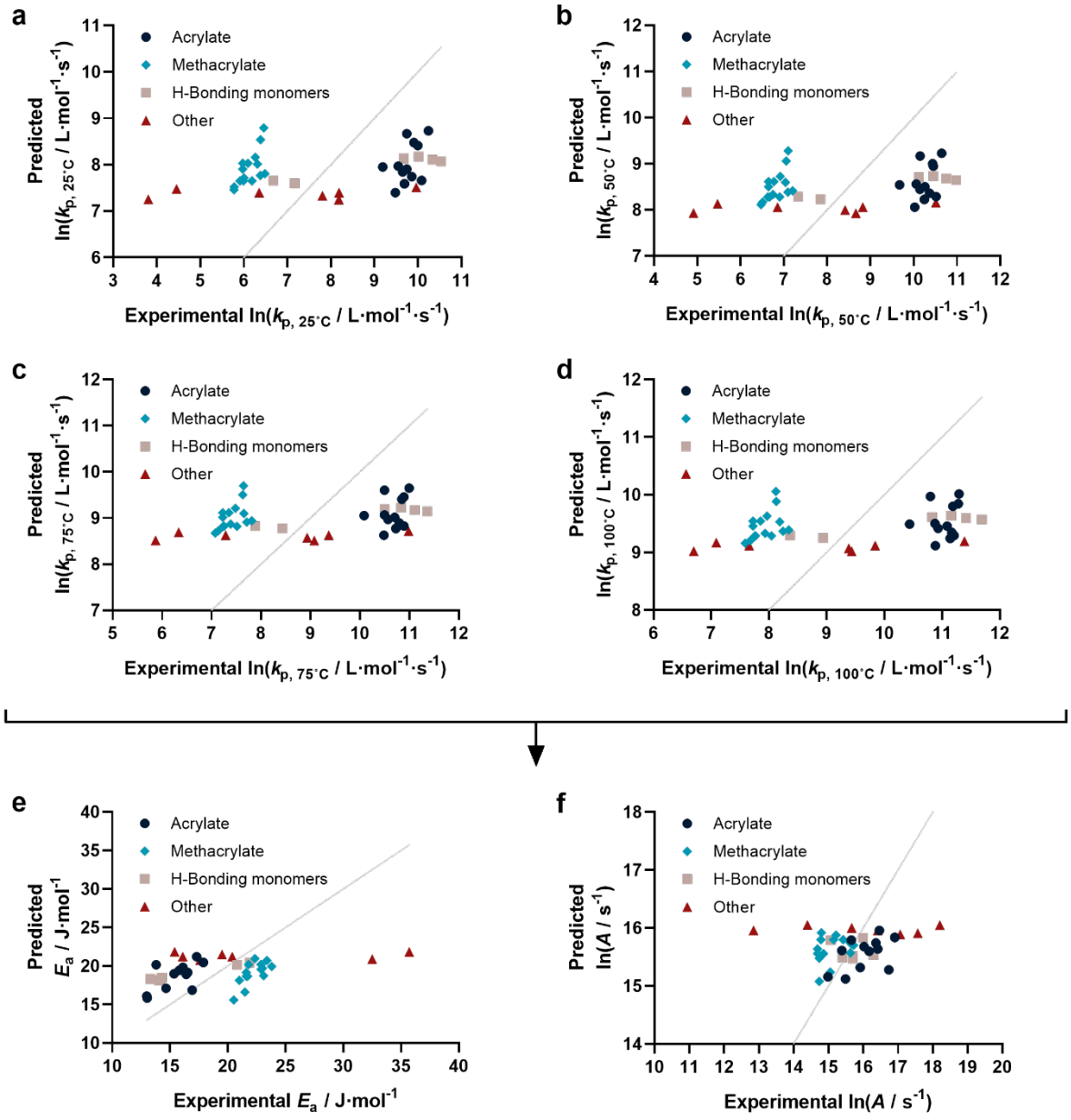


Figure 4-9: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 41) with the molecular weight as parameter. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	1.839	1.722	1.623	1.540	4.408	0.932
R2 Value	0.083	0.076	0.069	0.063	0.152	0.030
Variance Predictions	3.468	0.123	0.095	0.073	2.687	0.065
Variance Residuals	1.722	3.039	2.700	2.431	19.912	0.891

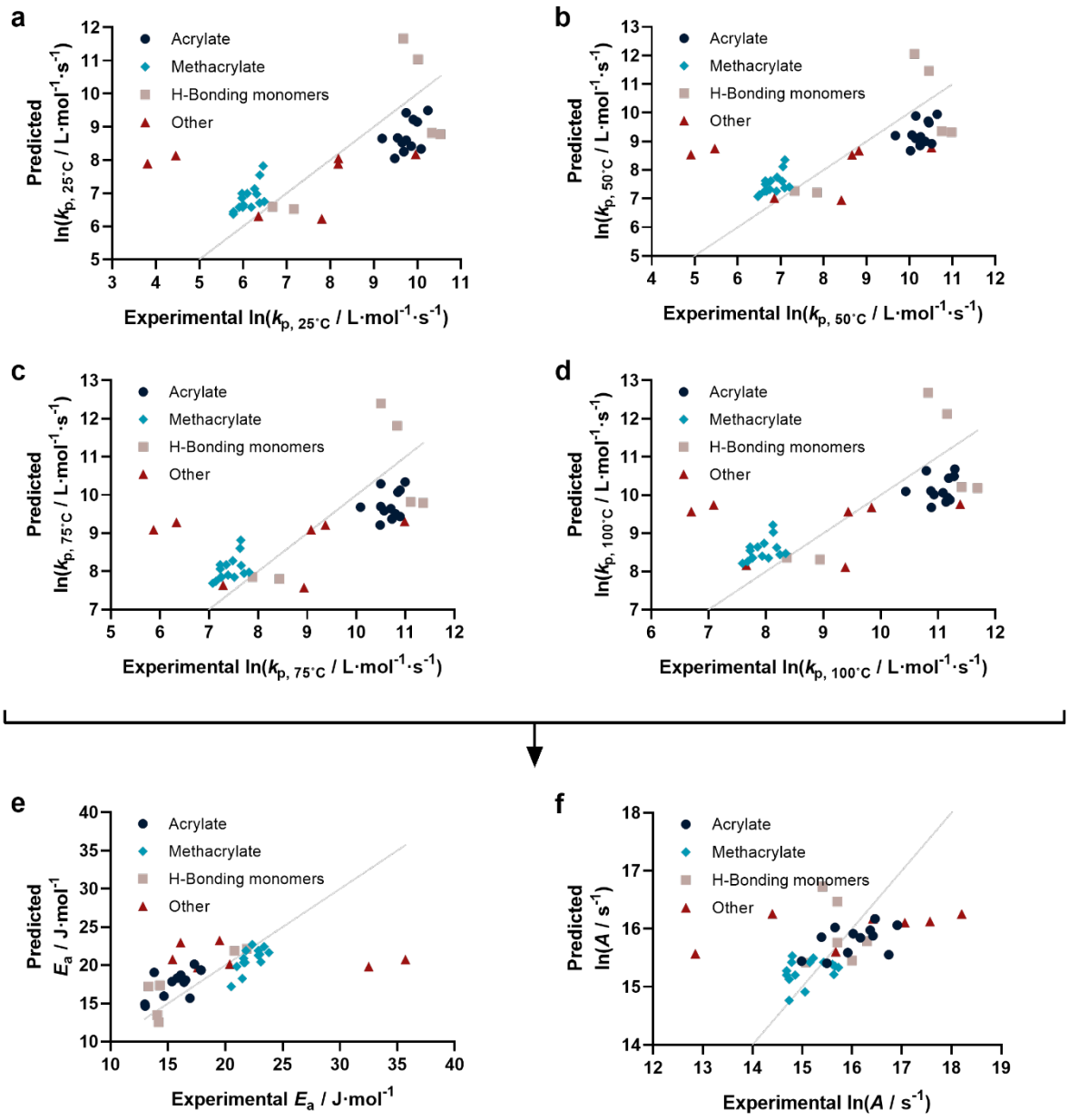


Figure 4-10: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 41) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers) as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	1.334	1.232	1.146	1.076	3.914	0.817
R2 Value	0.518	0.527	0.536	0.543	0.332	0.255
Variance Predictions	1.561	1.377	1.228	1.106	6.943	0.183
Variance Residuals	1.825	1.555	1.348	1.186	15.703	0.684

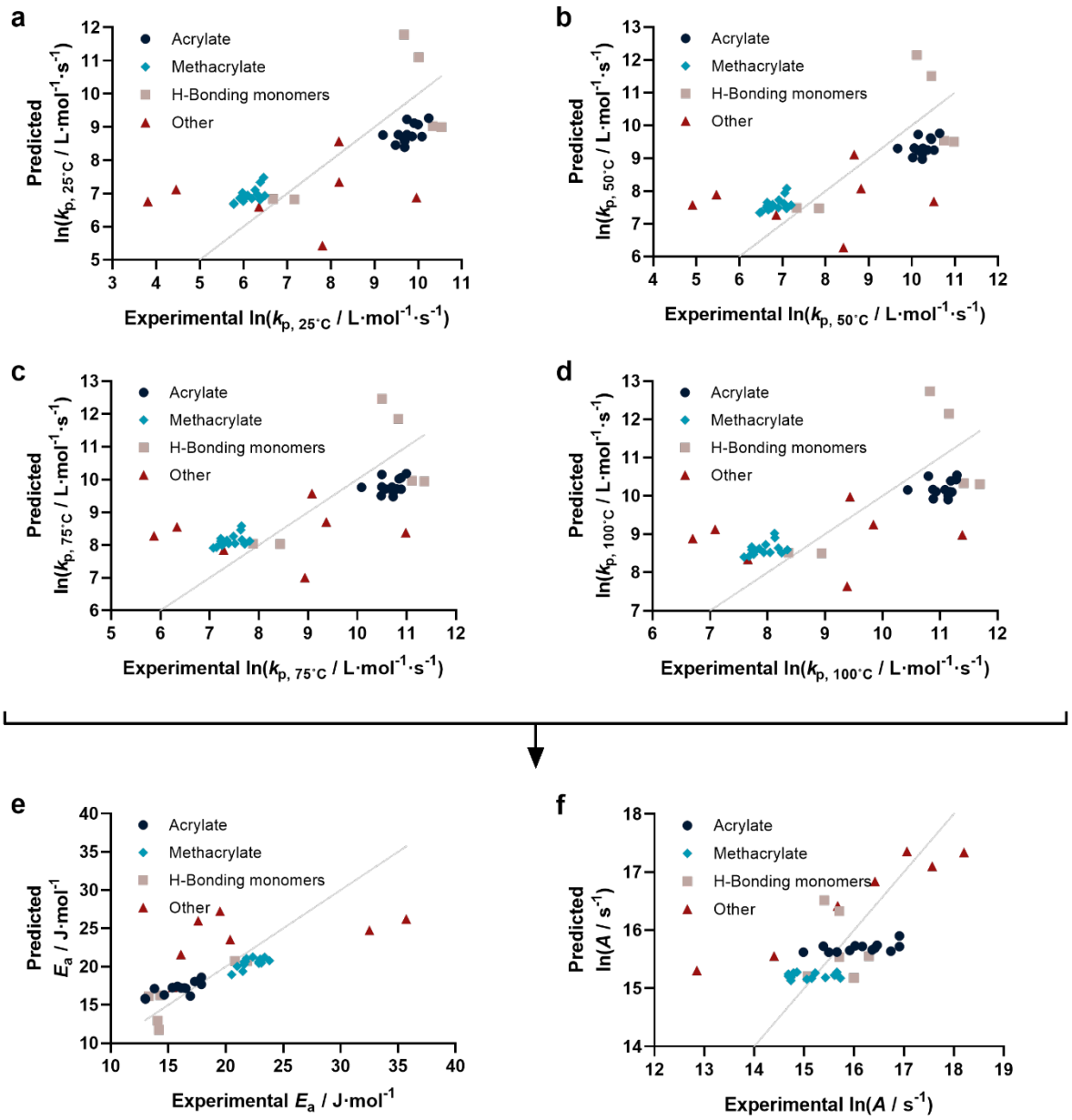


Figure 4-11: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 41) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers) and inductive effect of the tail group as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	1.256	1.171	1.100	1.041	3.271	0.708
R2 Value	0.572	0.572	0.572	0.572	0.533	0.441
Variance Predictions	1.647	1.424	1.249	1.111	10.839	0.363
Variance Residuals	1.618	1.407	1.241	1.110	10.970	0.513

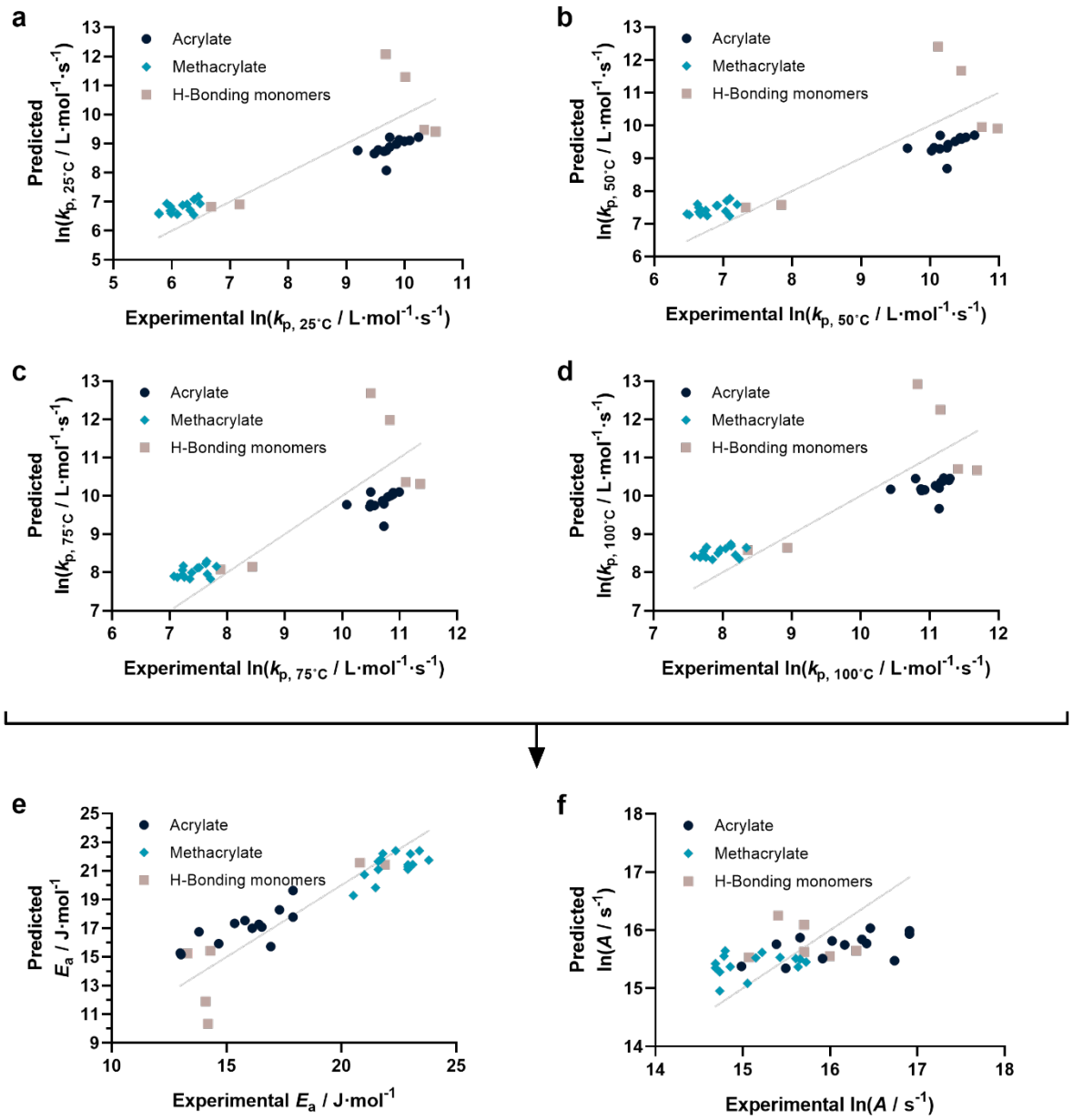


Figure 4-12: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all acrylates and methacrylates (n = 34) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers) and inductive effect of the tail group as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.899	0.861	0.829	0.803	1.546	0.545
R2 Value	0.761	0.753	0.745	0.736	0.822	0.301
Variance Predictions	1.998	1.740	1.532	1.357	10.012	0.078
Variance Residuals	0.834	0.764	0.708	0.664	2.461	0.306

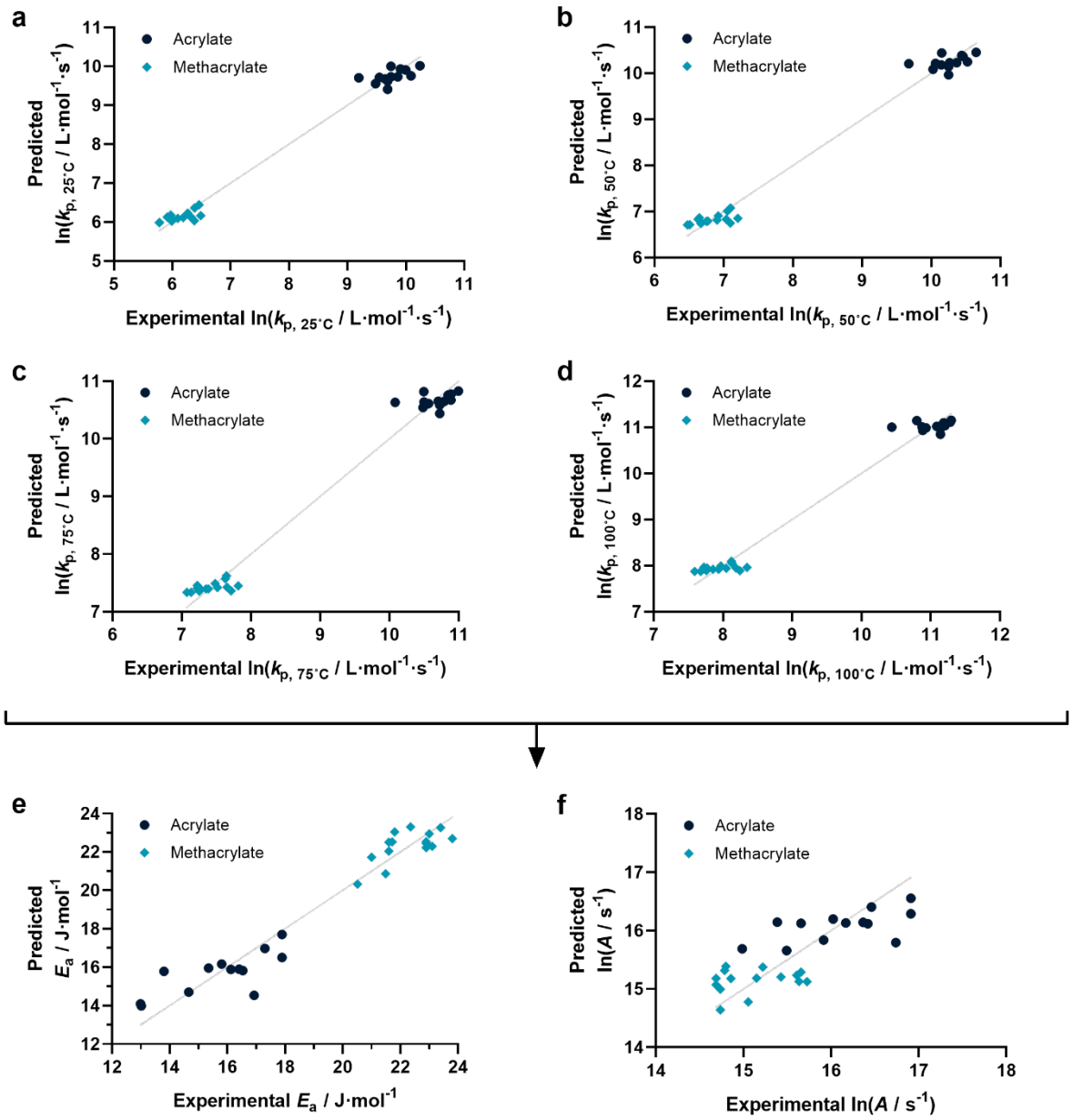


Figure 4-13: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all acrylates and methacrylates without H-bonding monomers (n = 28) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers) and inductive effect of the tail group as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.198	0.202	0.209	0.216	0.907	0.429
R2 Value	0.988	0.986	0.984	0.981	0.933	0.618
Variance Predictions	3.273	2.905	2.607	2.362	12.203	0.284
Variance Residuals	0.041	0.042	0.045	0.049	0.853	0.190

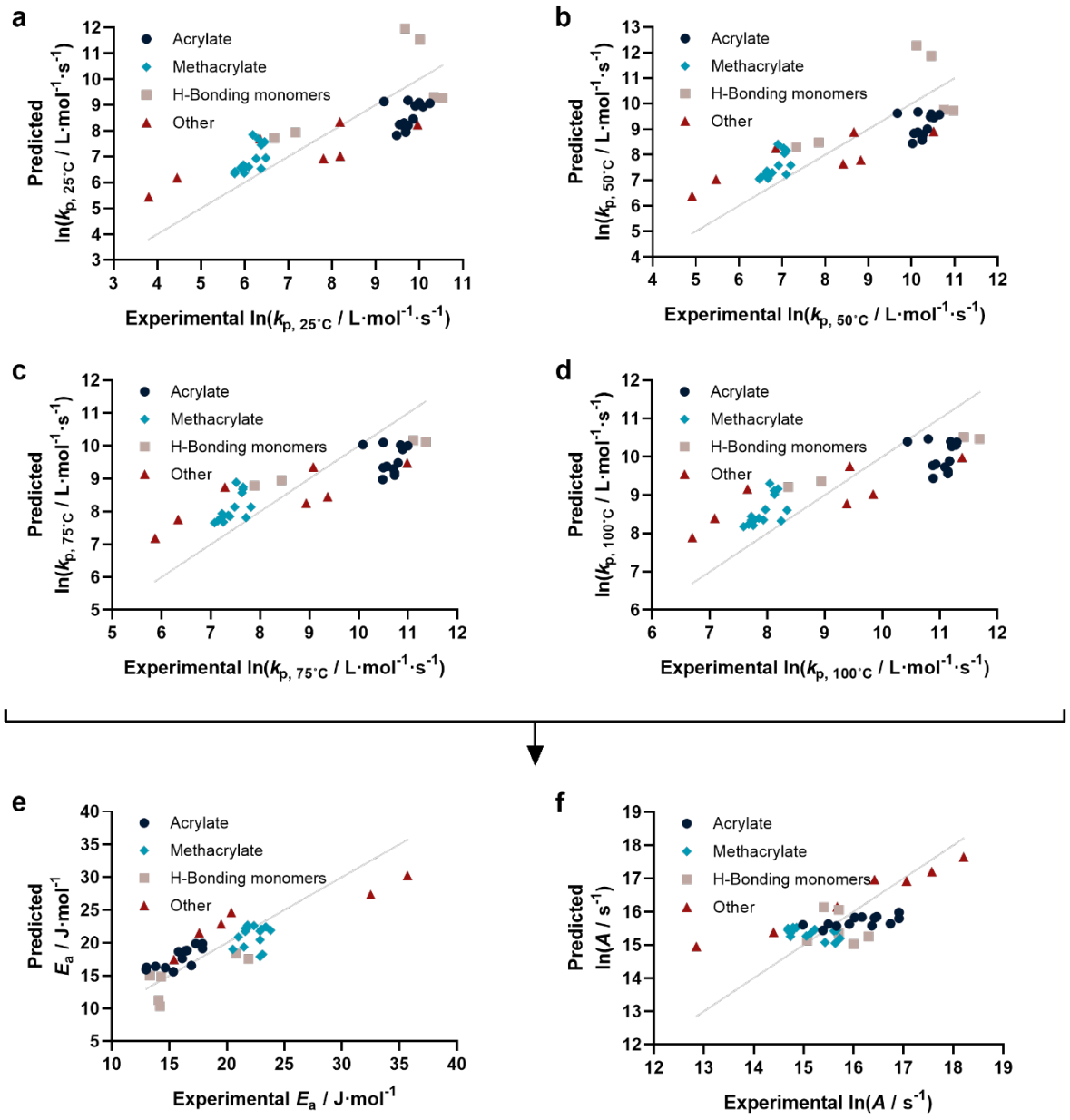


Figure 4-14: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 41) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, dipole moment, boiling point, melting point and Gibbs Free Energy as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	1.167	1.100	1.044	0.998	2.667	0.690
R2 Value	0.631	0.623	0.615	0.607	0.690	0.487
Variance Predictions	1.775	1.504	1.292	1.124	14.206	0.345
Variance Residuals	1.396	1.241	1.117	1.020	7.289	0.471

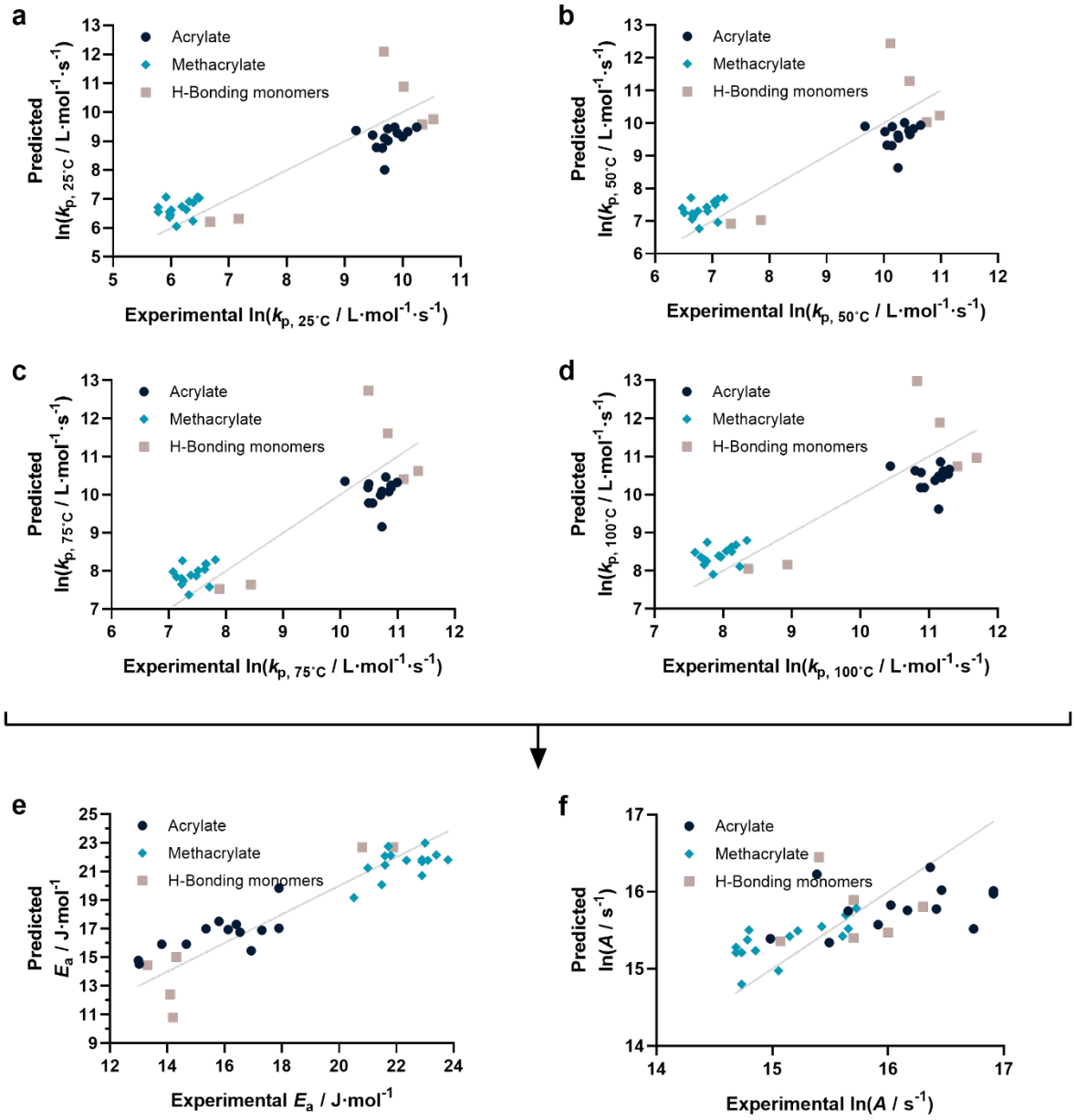


Figure 4-15: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all acrylates and methacrylates (n = 34) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, dipole moment, boiling point, melting point and Gibbs Free Energy as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.806	0.773	0.745	0.722	1.397	0.512
R2 Value	0.808	0.801	0.794	0.787	0.855	0.384
Variance Predictions	2.434	2.125	1.877	1.677	11.592	0.132
Variance Residuals	0.670	0.615	0.572	0.537	2.012	0.270

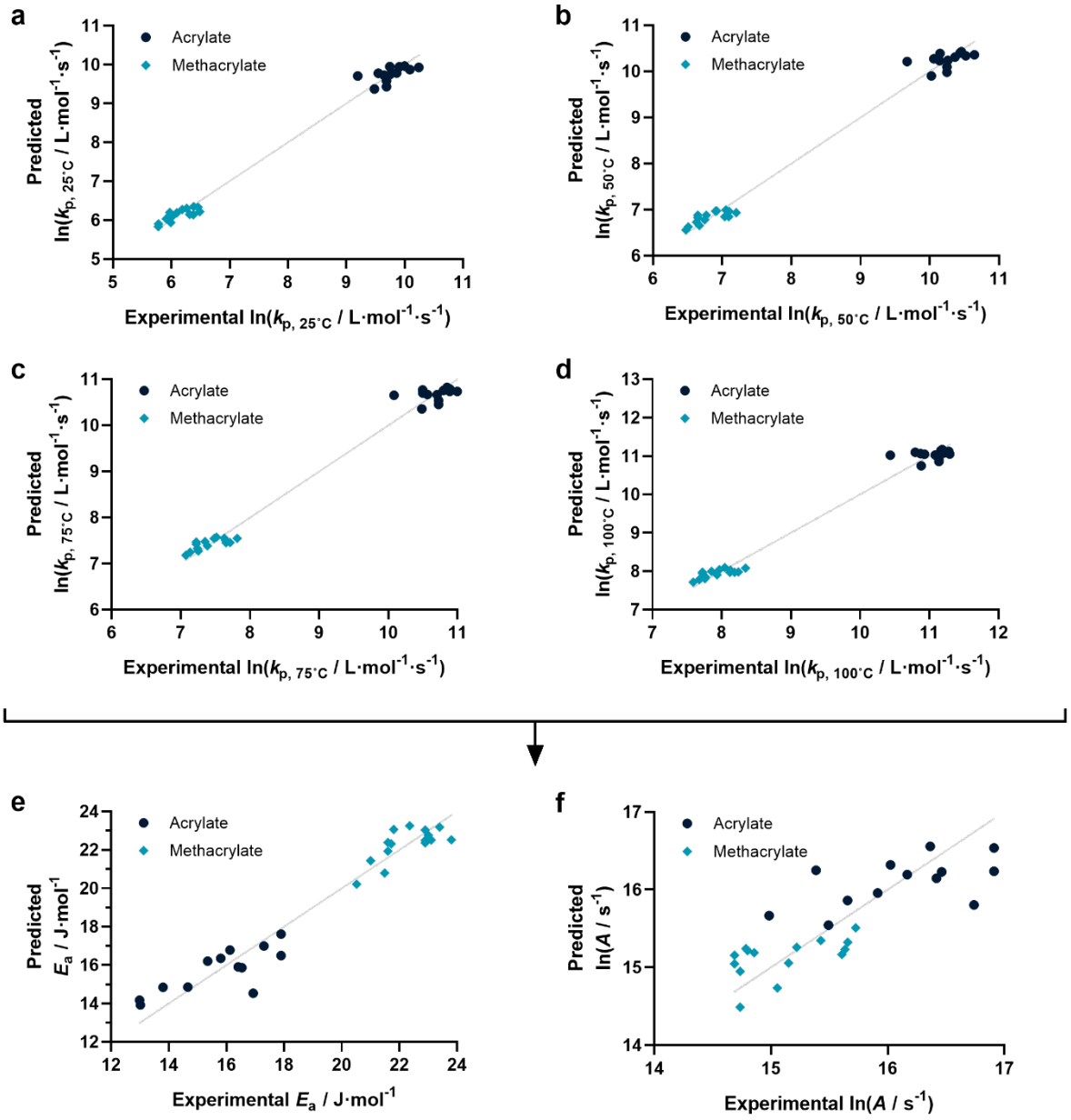


Figure 4-16: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all acrylates and methacrylates without H-bonding monomers (n = 28) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, dipole moment, boiling point, melting point and Gibbs Free Energy as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.181	0.186	0.192	0.200	0.849	0.403
R2 Value	0.990	0.988	0.986	0.984	0.941	0.661
Variance Predictions	3.277	2.915	2.621	2.376	12.065	0.320
Variance Residuals	0.034	0.036	0.038	0.042	0.748	0.169

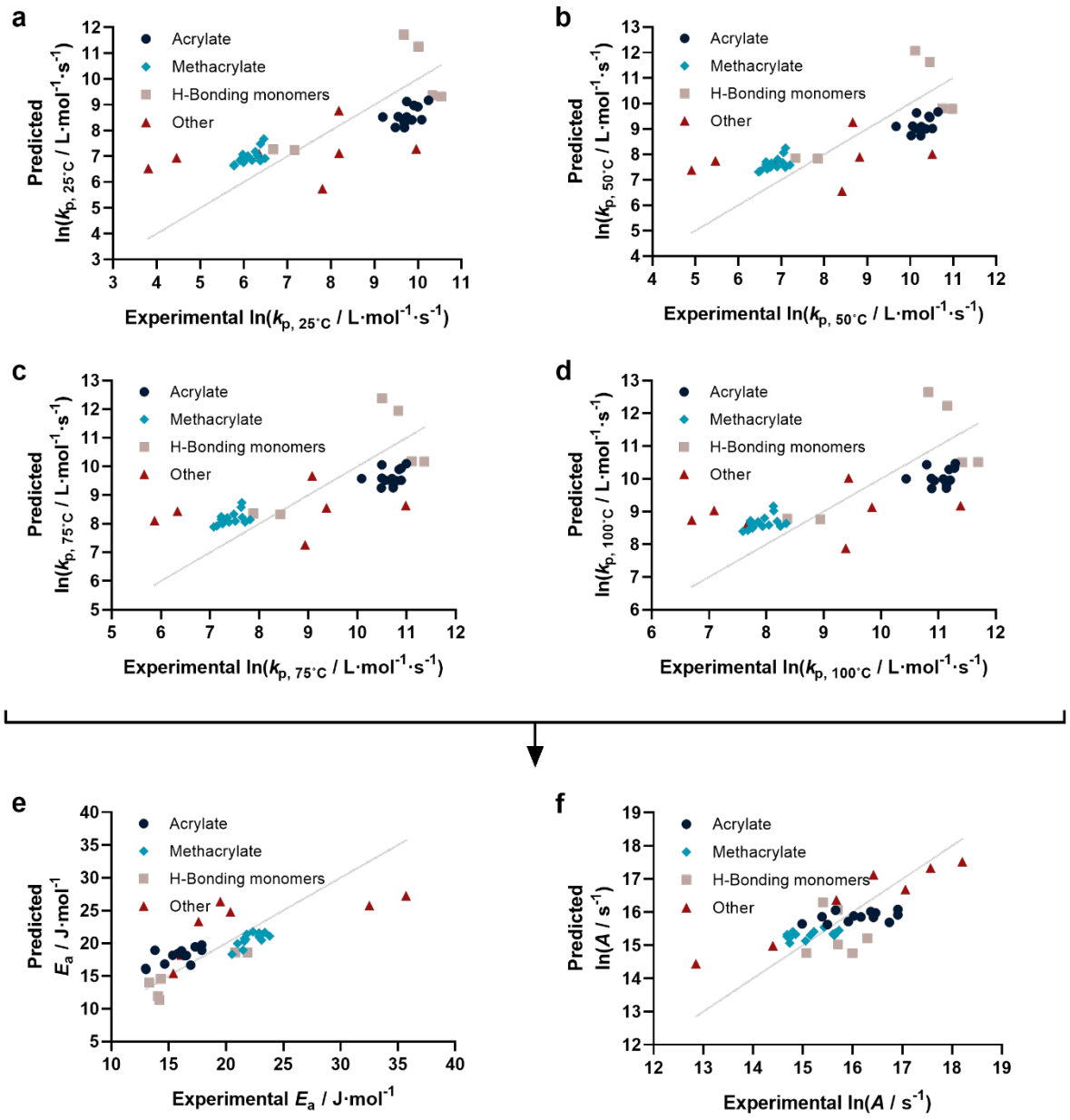


Figure 4-17: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 41) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, the effect of H-donors and H-acceptors as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	1.272	1.189	1.120	1.061	3.053	0.617
R2 Value	0.561	0.559	0.557	0.555	0.593	0.574
Variance Predictions	1.509	1.287	1.126	0.994	11.659	0.434
Variance Residuals	1.658	1.449	1.285	1.154	9.554	0.391

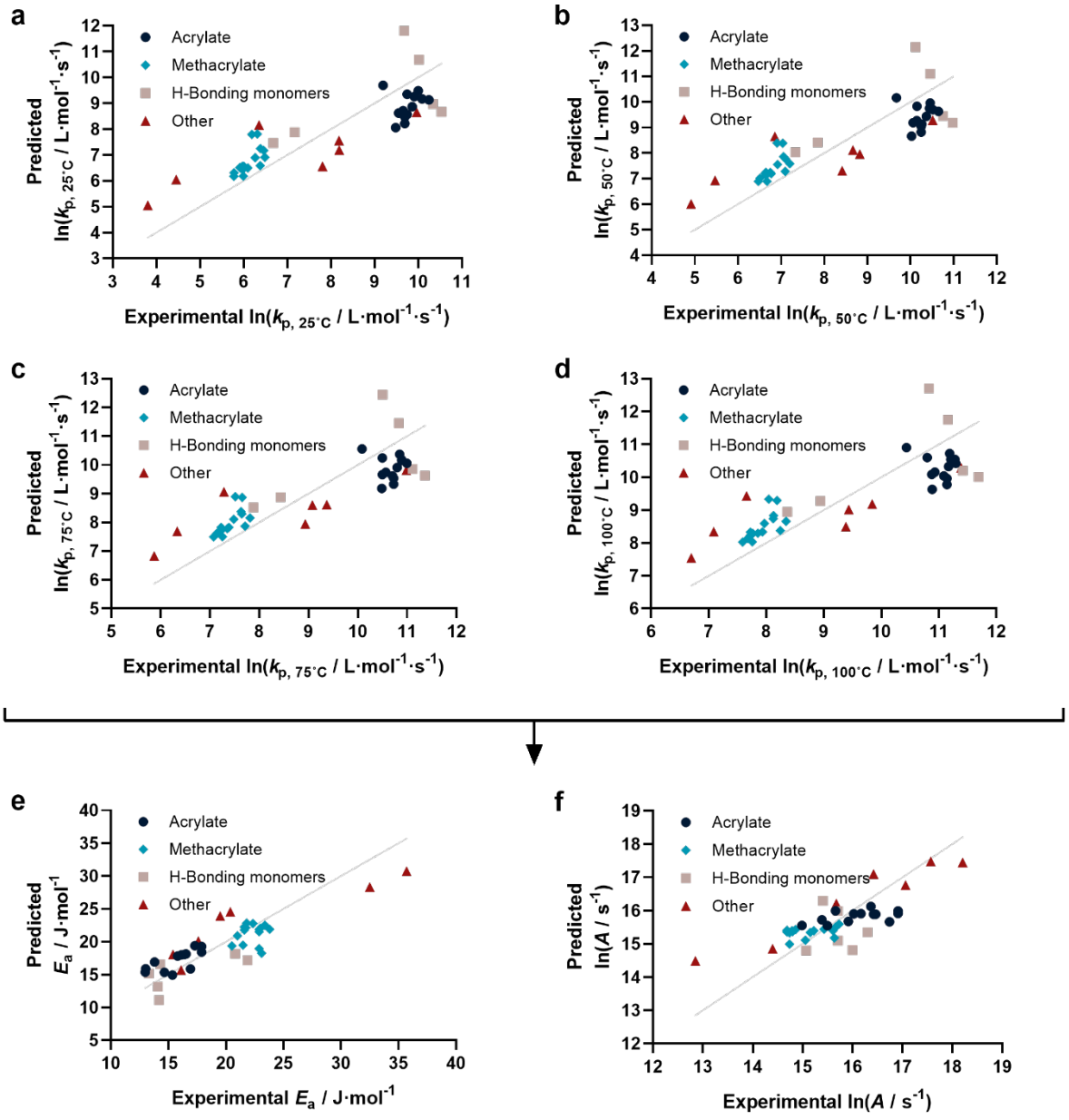


Figure 4-18: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 41) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, dipole moment, boiling point, melting point, Gibbs Free Energy, the effect of H-donors and H-acceptors as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	1.048	0.986	0.936	0.893	2.421	0.606
R2 Value	0.703	0.697	0.691	0.685	0.744	0.590
Variance Predictions	1.901	1.622	1.402	1.230	14.638	0.426
Variance Residuals	1.125	0.998	0.898	0.818	6.006	0.377

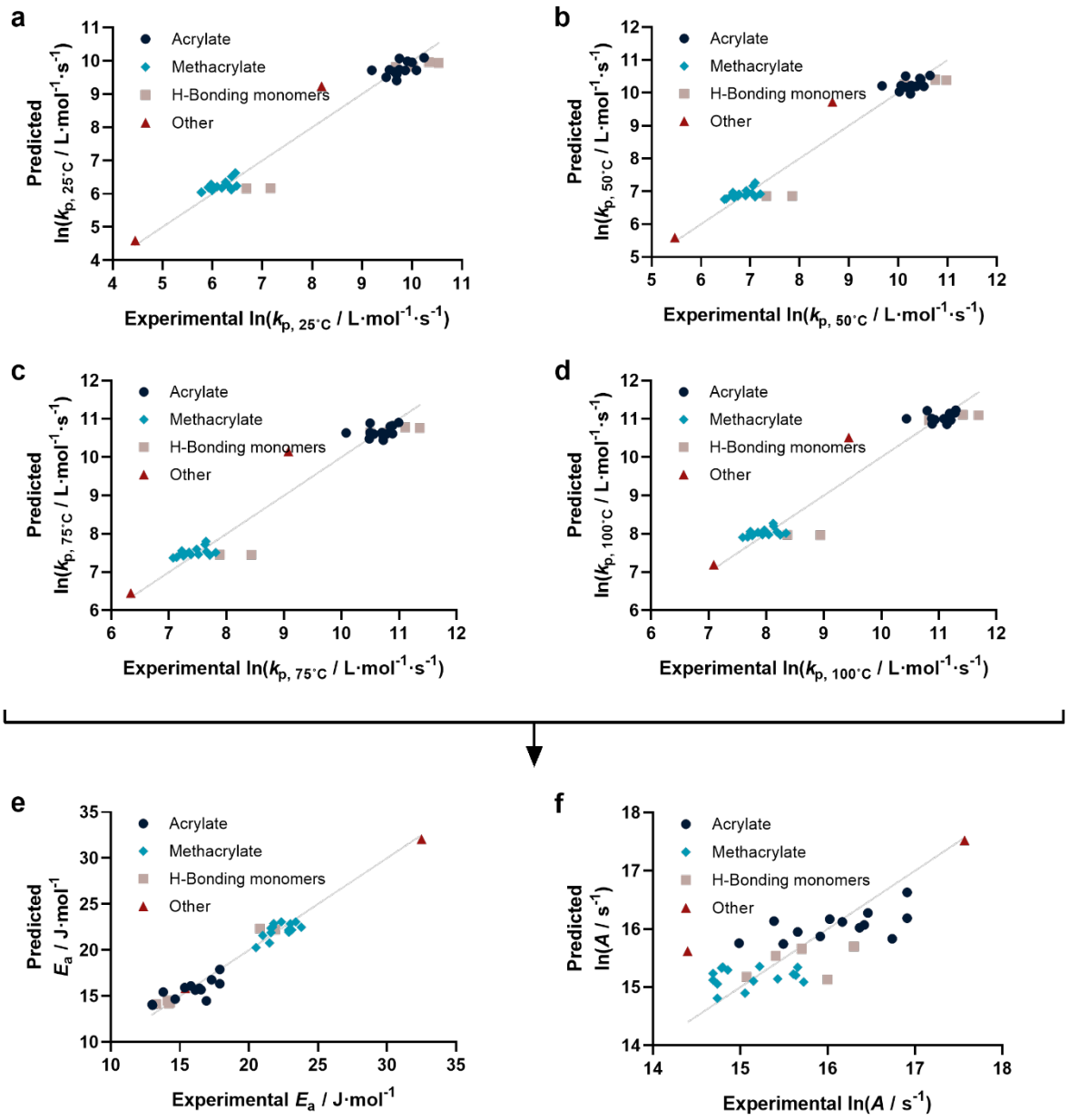


Figure 4-19: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 36) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group and the dissociation constants as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.343	0.341	0.342	0.344	0.874	0.472
R2 Value	0.967	0.963	0.957	0.952	0.958	0.595
Variance Predictions	3.383	2.947	2.599	2.318	17.495	0.305
Variance Residuals	0.121	0.120	0.121	0.122	0.786	0.229

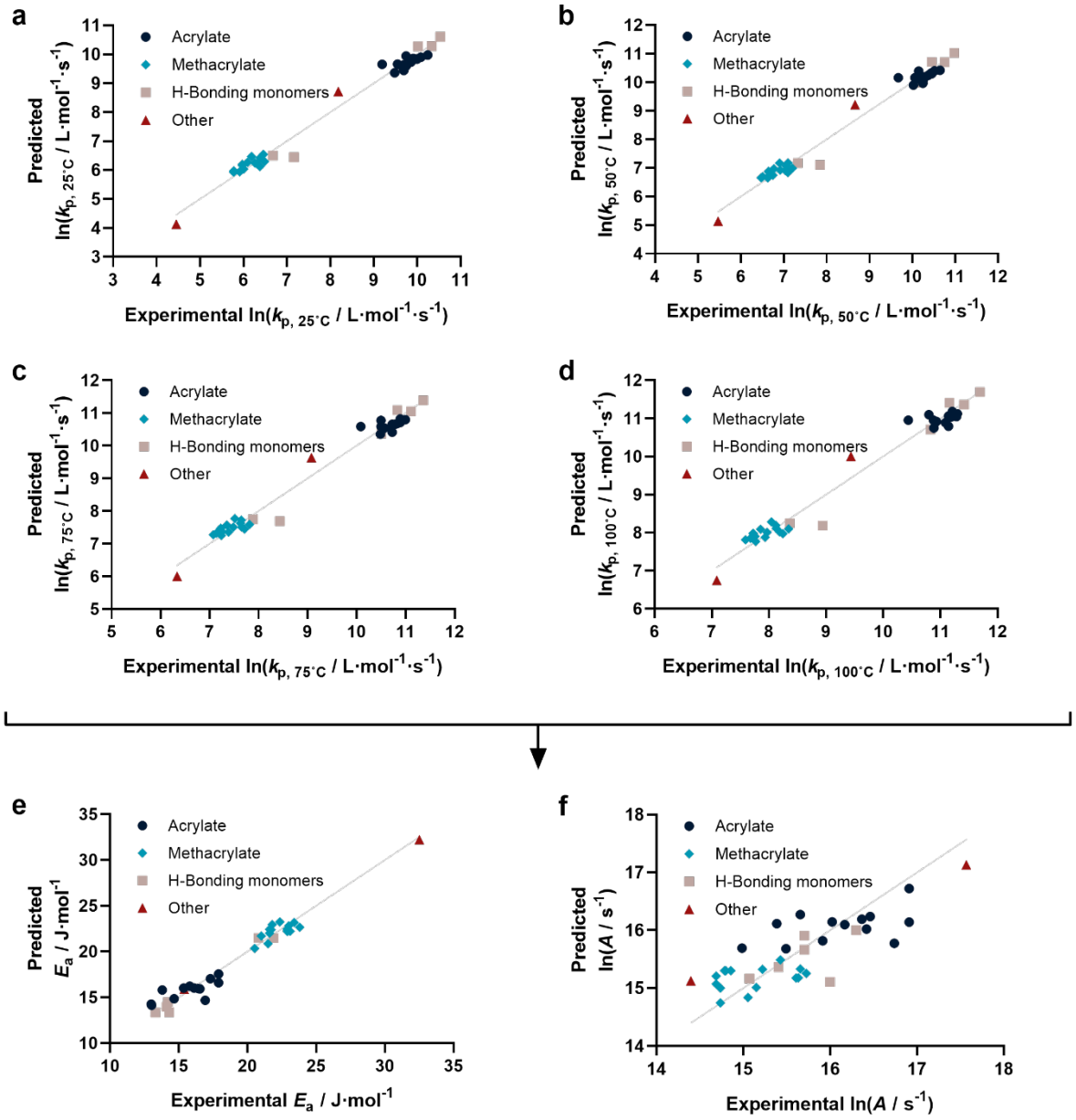


Figure 4-20: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 36) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, the effect of H-donors and H-acceptors, and the dissociation constants as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.232	0.238	0.245	0.253	0.824	0.439
R2 Value	0.985	0.982	0.978	0.974	0.963	0.649
Variance Predictions	3.540	3.088	2.727	2.435	17.788	0.300
Variance Residuals	0.055	0.058	0.062	0.066	0.698	0.198

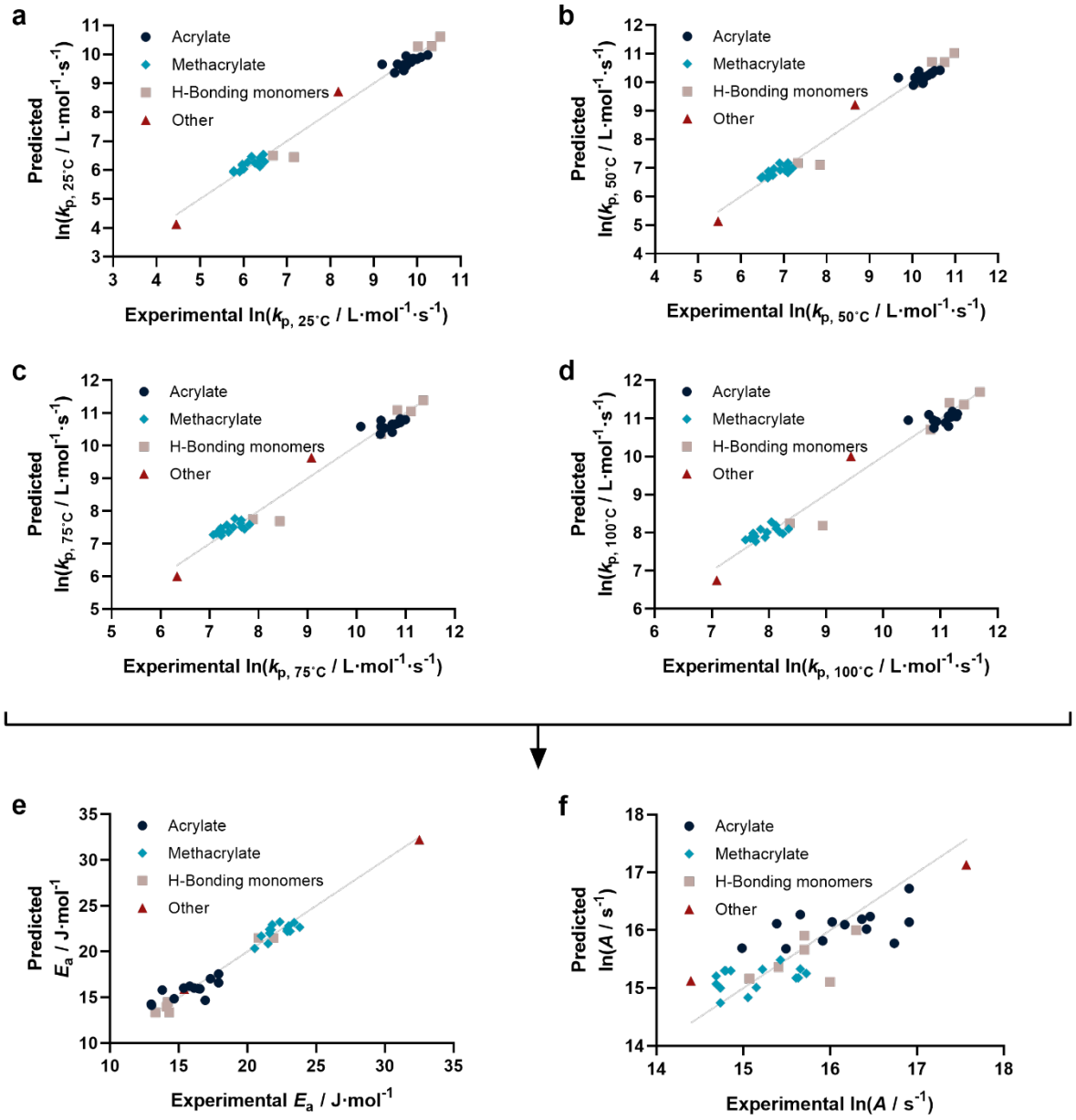


Figure 4-21: Correlation plots of the natural logarithm of the predicted propagation rate constant ($\ln(k_p)$) versus the experimental $\ln(k_p)$ for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 36) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, dipole moment, boiling point, melting point, Gibbs Free Energy, the effect of H-donors and H-acceptors, and the dissociation constants as parameters. From the predicted $\ln(k_p)$ results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor ($\ln(A)$) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figuree	Figure f
RMSE	0.172	0.178	0.185	0.193	0.792	0.386
R2 Value	0.992	0.990	0.988	0.985	0.965	0.729
Variance Predictions	3.497	3.057	2.706	2.421	17.734	0.398
Variance Residuals	0.031	0.032	0.035	0.038	0.645	0.153

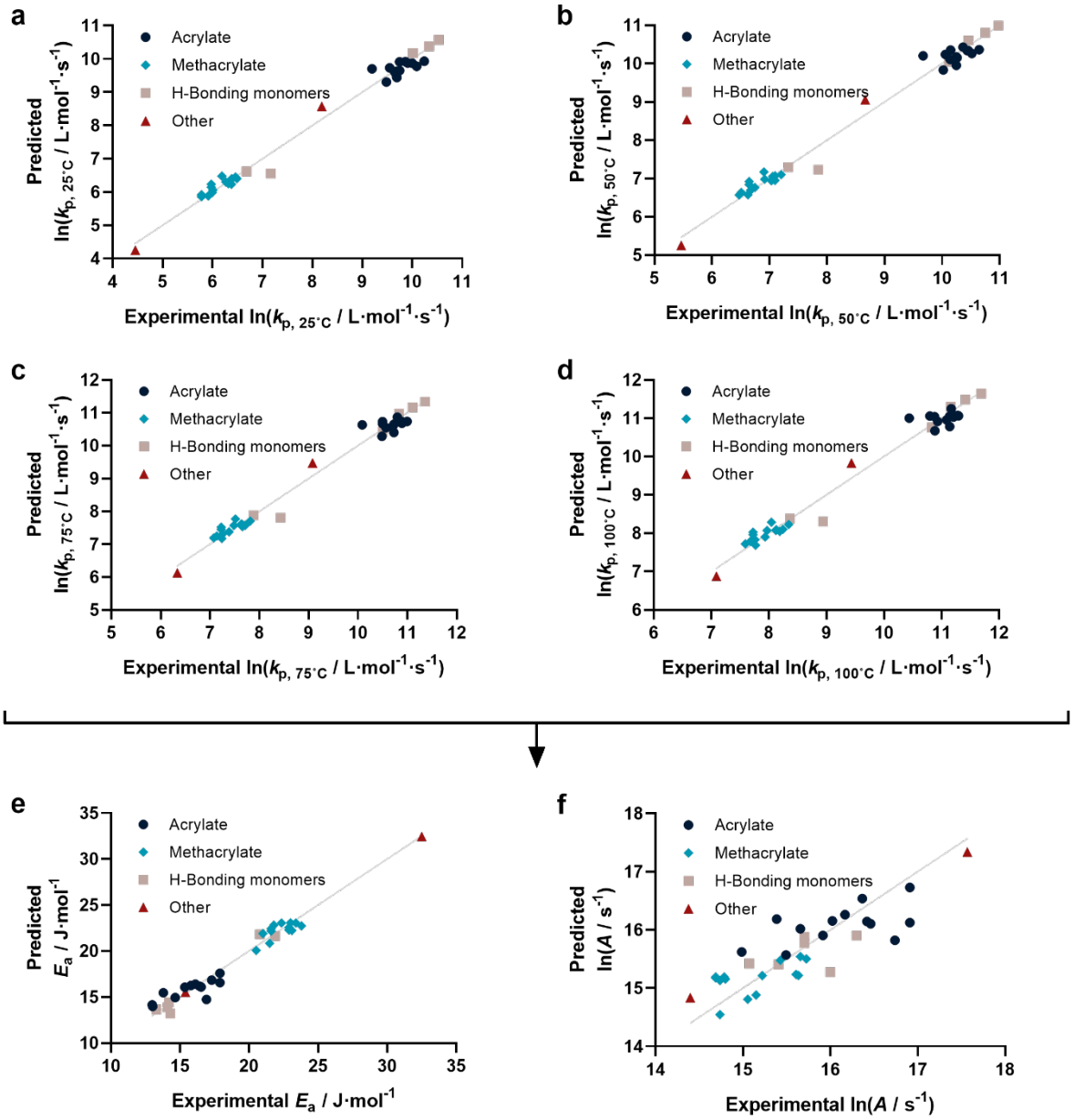


Figure 4-22: Correlation plots of the natural logarithm of the predicted propagation rate constant ($\ln(k_p)$) versus the experimental $\ln(k_p)$ for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 35) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, the effect of H-donors and H-acceptors, the dissociation constants as parameters, predicted boiling point, predicted refractive index and predicted density as parameters. From the predicted $\ln(k_p)$ results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor ($\ln(A)$) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.207	0.210	0.216	0.222	0.907	0.429
R2 Value	0.987	0.986	0.983	0.980	0.933	0.618
Variance Predictions	3.532	3.079	2.718	2.426	12.206	0.284
Variance Residuals	0.044	0.045	0.048	0.051	0.853	0.190

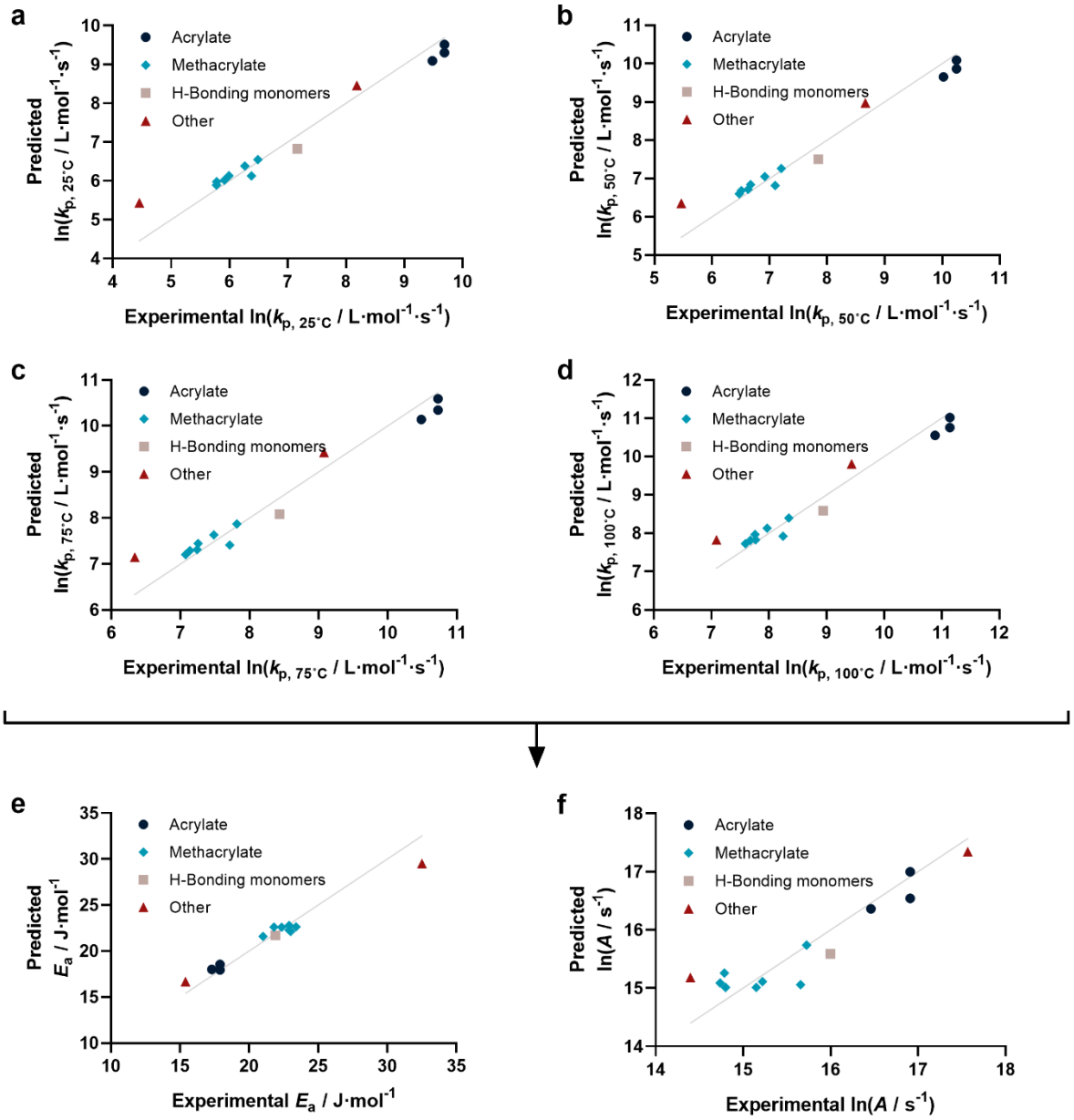


Figure 4-23: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 13) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, the effect of H-donors and H-acceptors, the dissociation constants, experimental boiling point, experimental refractive index and experimental density as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.355	0.339	0.325	0.316	1.050	0.369
R2 Value	0.953	0.951	0.950	0.948	0.933	0.851
Variance Predictions	1.991	1.769	1.600	1.459	10.617	0.673
Variance Residuals	0.136	0.123	0.114	0.107	1.186	0.148

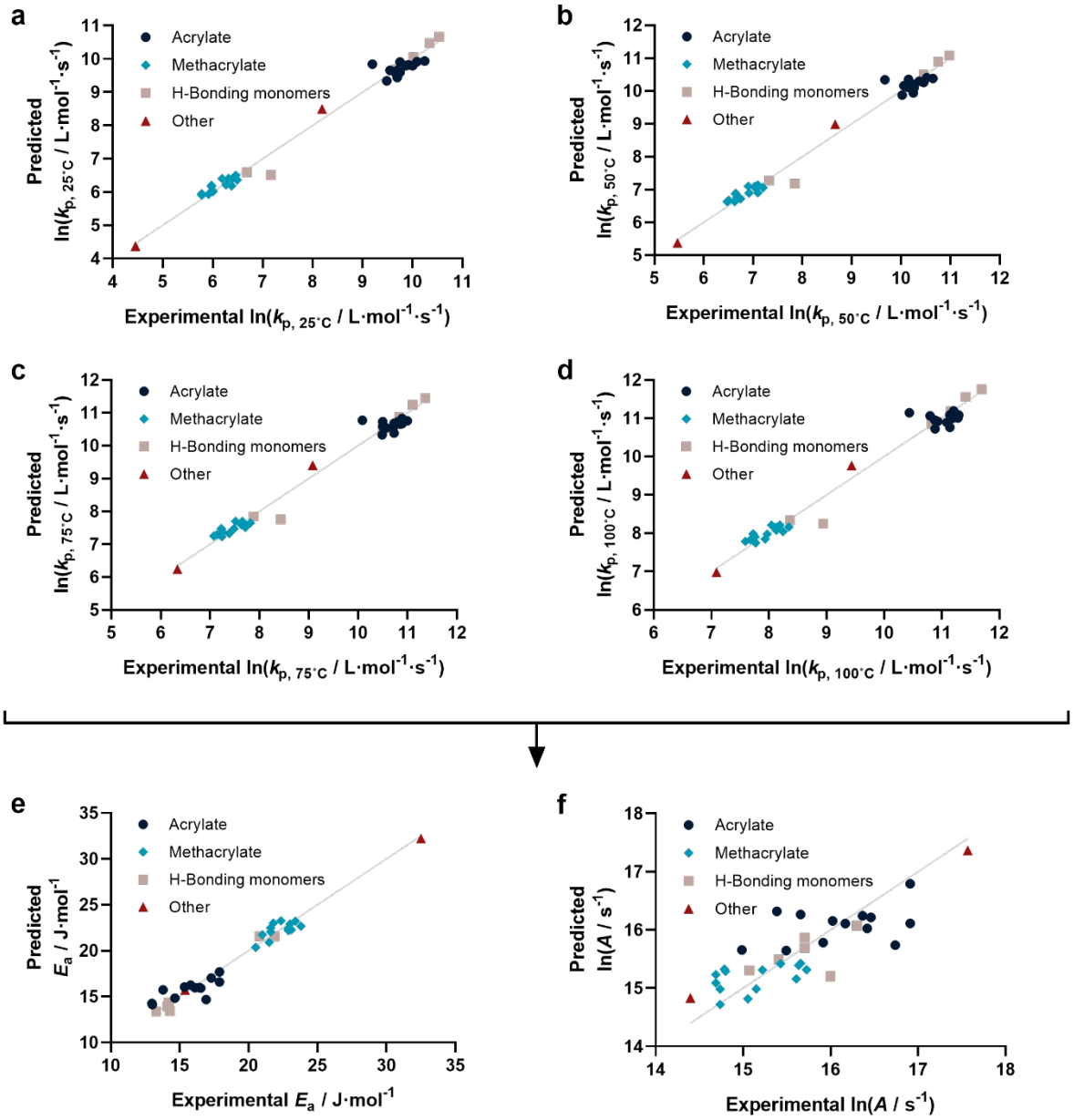


Figure 4-24: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 35) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, the effect of H-donors and H-acceptors, the dissociation constants and ACDlogP as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.212	0..218	0.225	0.233	0.823	0.423
R2 Value	0.987	0.985	0.982	0.978	0.963	0.674
Variance Predictions	3.512	3.062	2.704	2.413	18.061	0.343
Variance Residuals	0.046	0.049	0.052	0.056	0.698	0.184

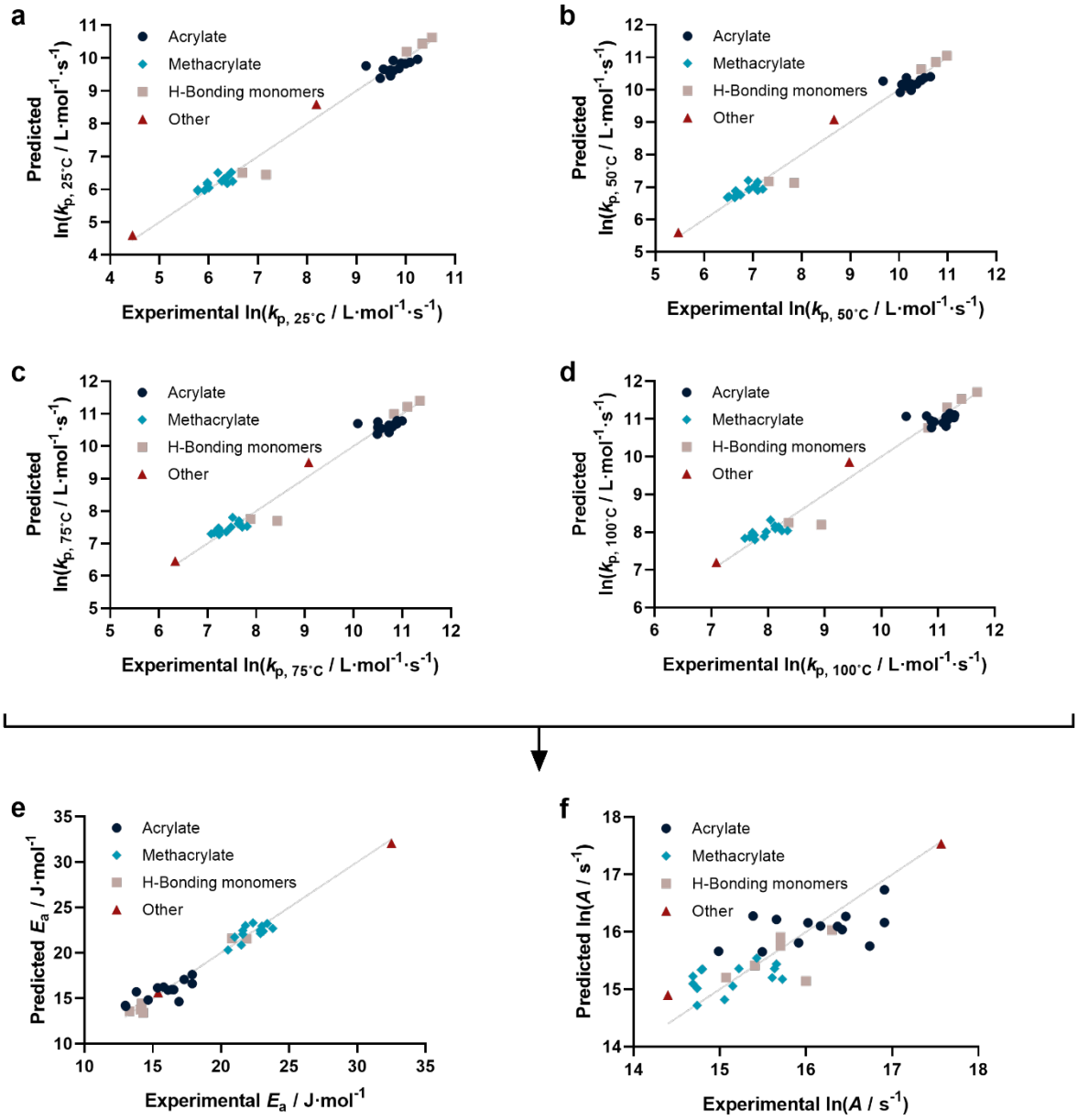


Figure 4-25: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 35) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, the effect of H-donors and H-acceptors, the dissociation constants and polarizability as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.227	0.232	0.238	0.245	0.823	0.427
R2 Value	0.985	0.983	0.979	0.976	0.963	0.666
Variance Predictions	3.454	3.010	2.658	2.368	18.064	0.349
Variance Residuals	0.053	0.055	0.058	0.062	0.701	0.188

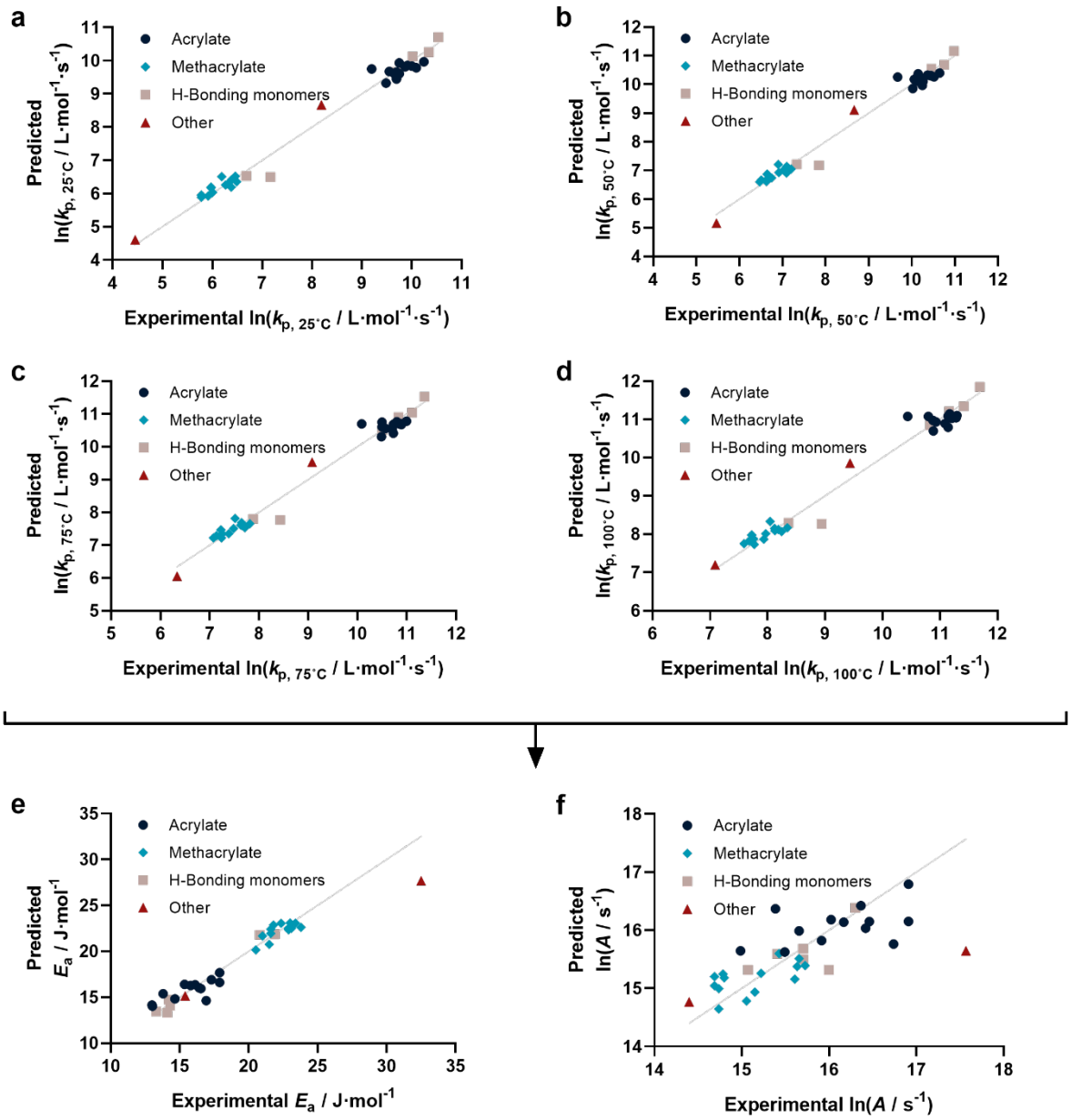


Figure 4-26: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 35) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, the effect of H-donors and H-acceptors, the dissociation constants and surface tension as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.222	0.226	0.231	0.236	1.139	0.512
R2 Value	0.986	0.983	0.981	0.977	0.930	0.521
Variance Predictions	3.450	3.090	2.725	2.430	15.167	0.276
Variance Residuals	0.050	0.053	0.055	0.058	1.316	0.267

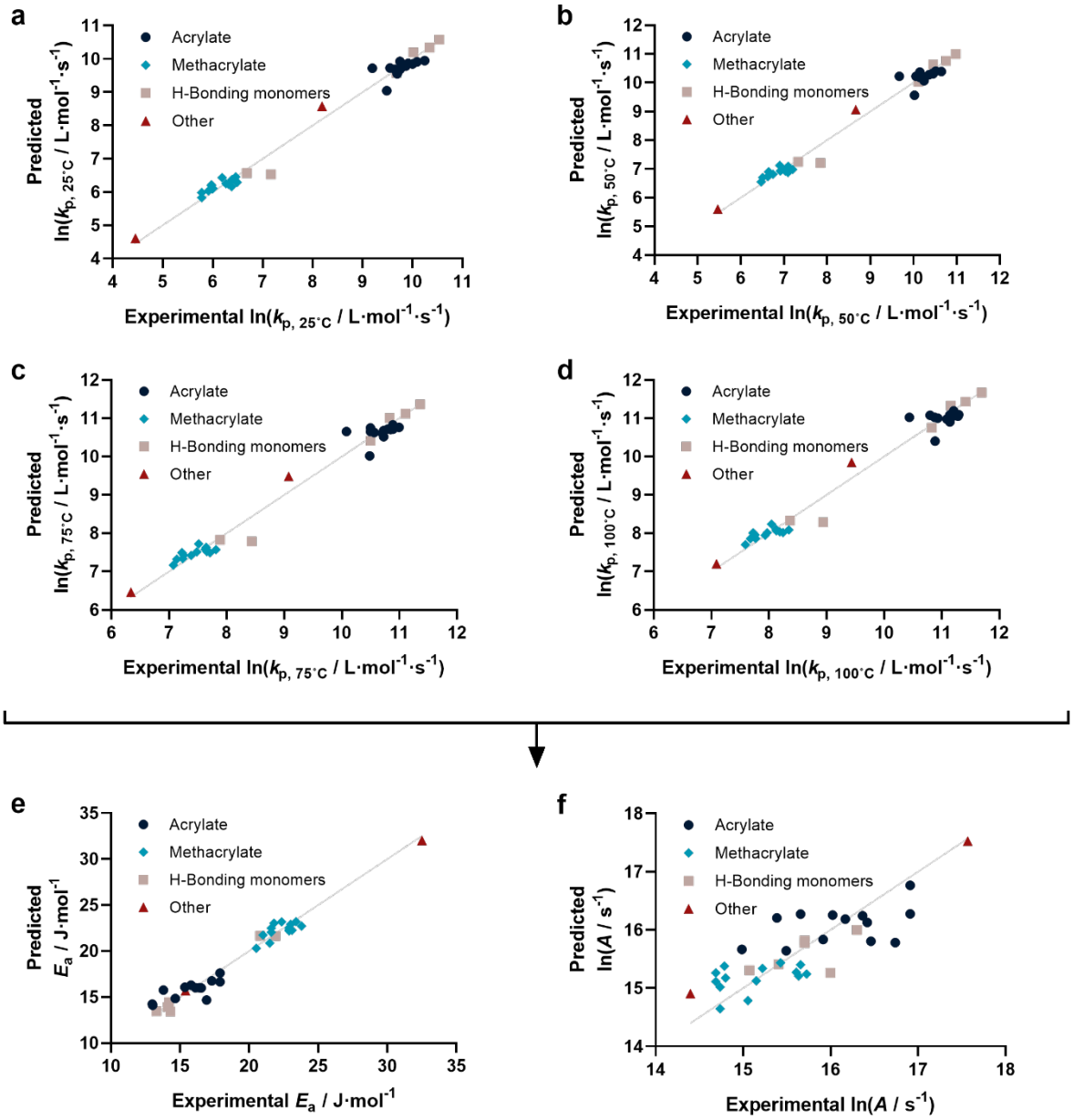


Figure 4-27: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 35) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, the effect of H-donors and H-acceptors, the dissociation constants and vapor pressure as parameters. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.219	0.224	0.230	0.237	0.825	0.421
R2 Value	0.987	0.984	0.981	0.977	0.963	0.676
Variance Predictions	3.455	3.012	2.660	2.374	17.927	0.355
Variance Residuals	0.049	0.052	0.054	0.058	0.701	0.183

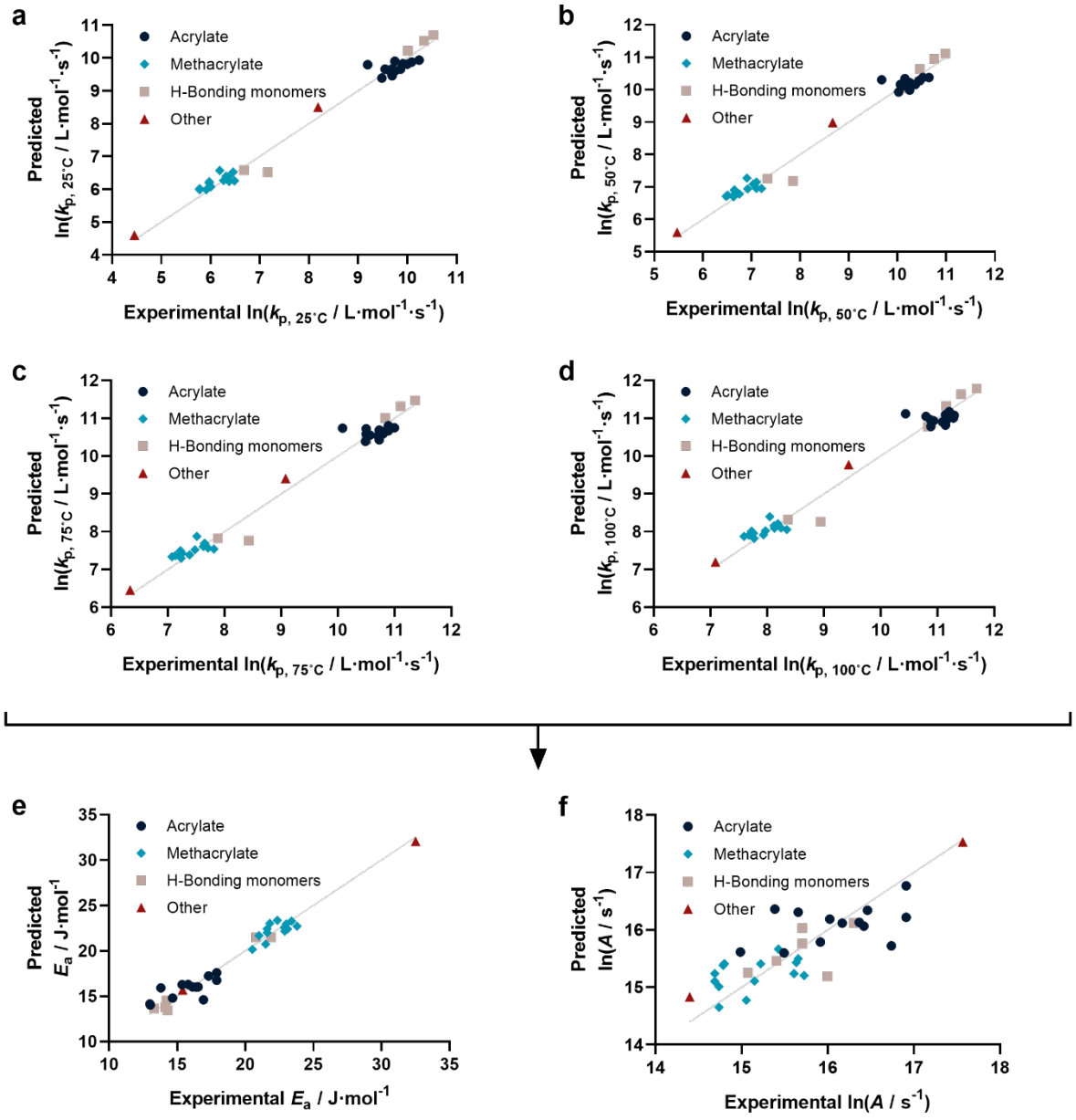


Figure 4-28: Correlation plots of the natural logarithm of the predicted propagation rate constant (ln(k_p)) versus the experimental ln(k_p) for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 35) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, the effect of H-donors and H-acceptors, the dissociation constants and polarizability as parameters. A selection of monomers (n = 13) is given a weight of 1 while the rest (n = 22) has a weight of 0.5. From the predicted ln(k_p) results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor (ln(A)) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	0.228	0.233	0.240	0.247	0.832	0.432
R2 Value	0.985	0.983	0.979	0.975	0.963	0.660
Variance Predictions	3.149	2.977	2.629	2.348	17.910	0.366
Variance Residuals	0.053	0.055	0.058	0.062	0.713	0.191

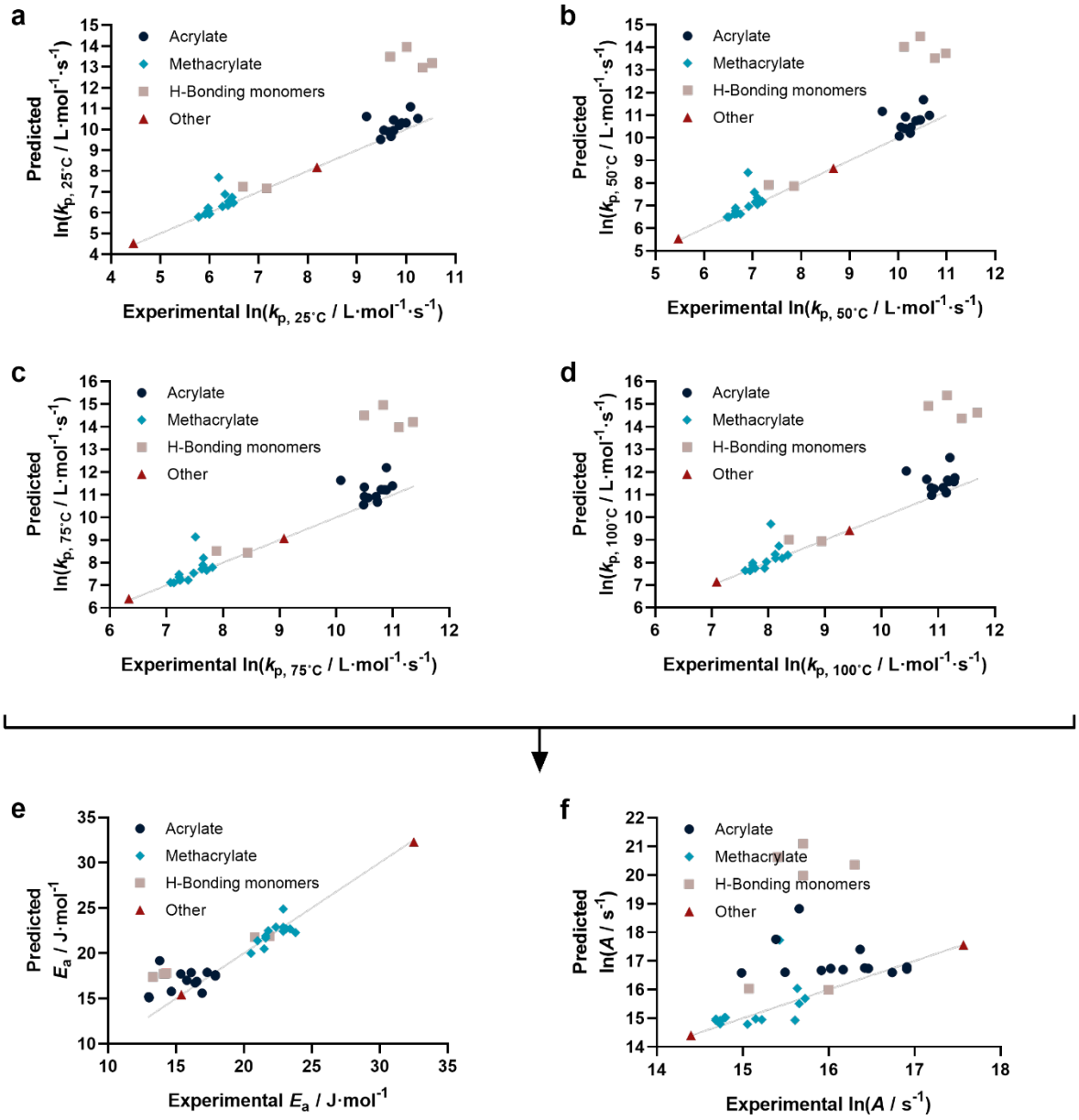


Figure 4-29: Correlation plots of the natural logarithm of the predicted propagation rate constant ($\ln(k_p)$) versus the experimental $\ln(k_p)$ for different temperatures: 25˚C, 50˚C, 75˚C and 100˚C for a, b, c and d respectively. The predictions were generated via a cross validated Ridge regression using all available data (n = 35) with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, the effect of H-donors and H-acceptors, the dissociation constants and polarizability as parameters. A selection of monomers (n = 13) is given a weight of 1 while the rest (n = 22) has a weight of 0.5. From the predicted $\ln(k_p)$ results, a prediction for the activation energy (E_a) and the natural logarithm of the pre-exponential factor ($\ln(A)$) was made using the Arrhenius plot and formula. The predictions were plotted against their experimental value (e and f). The metrics of each figure are annotated in the table below.

	Figure a	Figure b	Figure c	Figure d	Figure e	Figure f
RMSE	1.209	1.255	1.296	1.332	1.829	1.857
R2 Value	0.588	0.492	0.389	0.280	0.819	-5.294
Variance Predictions	6.356	5.959	5.637	5.374	12.590	3.149
Variance Residuals	1.109	1.190	1.266	1.337	2.700	2.594

Arrhenius plots of four selected monomers

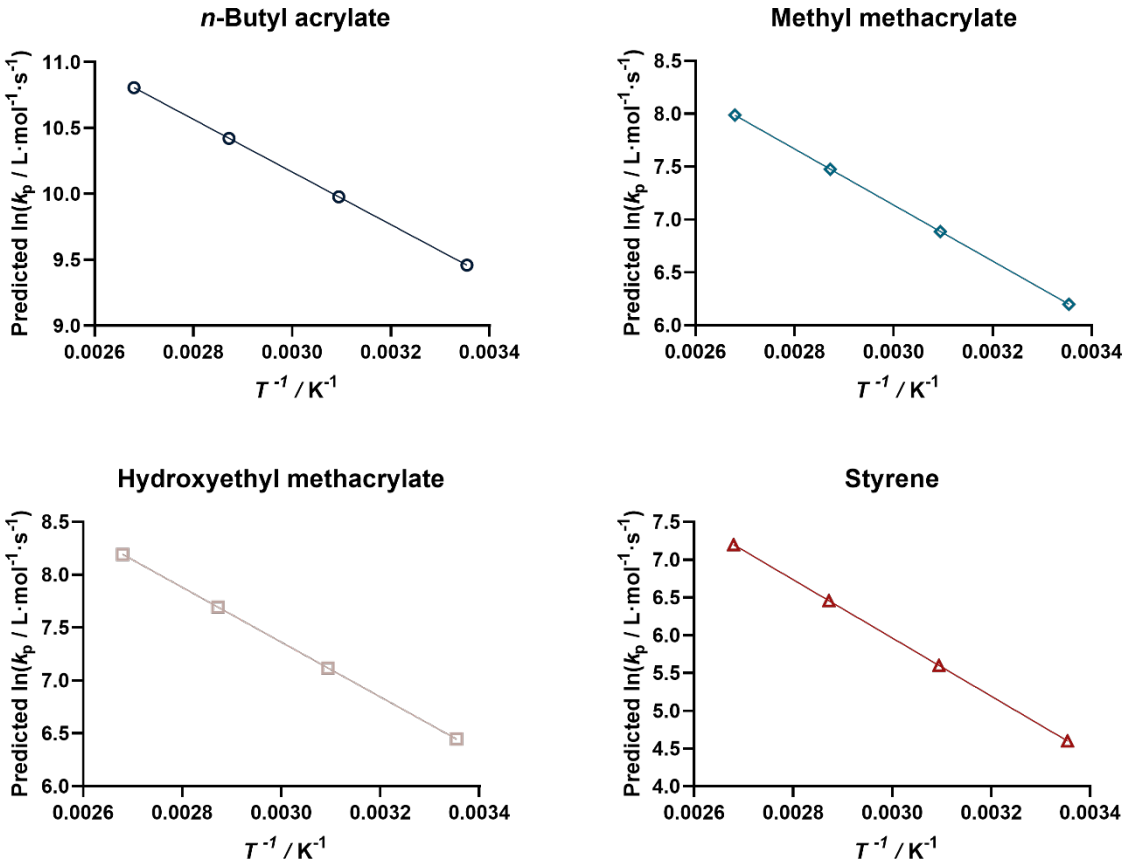


Figure 4-30: Arrhenius plots of n-Butyl acrylate, methyl methacrylate, hydroxyethyl methacrylate and styrene. The predicted ln(k_p) values were determined by a cross validated Ridge regression on all available ln(K_p) data with the molecular weight and a distinction between the type of monomer (blue = acrylate, teal = methacrylate, grey = H-bonding monomers and red = other monomers), inductive effect of the tail group, the effect of H-donors and H-acceptors, the dissociation constants and polarizability as parameters

Appendix B

1. Tables

Table B1: List wih monomers, their unique ID

Monomer name	Unique ID	SMILES	Mr
Acrylate p-biphenyl	1	C=CC(=O)OC1=CC=CC=C1C2=CC=CC=C2	224.25
(-) Menthyl acrylate	2	CC1CCC(C(C1)OC(=O)C=C)C(C)C	210.31
N-Cyanocinnamate ethyl	3	CCOC(=O)C(=CC1=CC=CC=C1)C#N	201.22
1,3-Butadiene, 1-acetoxy	4	CC(=O)OC=CC=C	112.13
2-Vinylbenzofuran	5	C=CC1=CC2=CC=CC=C2O1	144.17
Acenaphthylene	6	C1=CC2=C3C(=C1)C=CC3=CC=C2	152.19
Acrolein	7	C=CC=O	56.06
Acrylamide	8	C=CC(=O)N	71.08
Acrylamide N-octadecyl	9	CCCCCCCCCCCCCCCCCCNC(=O)C=C	323.60
Acrylate 2,4,6-tribromophenyl	10	C=CC(=O)OC1=C(C=C(C=C1Br)Br)Br	384.85
Acrylate 2-ethylhexyl	11	CCCCC(CC)COC(=O)C=C	184.28
Acrylate butyl	12	CCCCOC(=O)C=C	128.17
Acrylate cyclohexyl	13	C=CC(=O)OC1CCCCC1	154.21
Acrylate ethyl	14	CCOC(=O)C=C	100.12
Acrylate glycidyl	15	C=CC(=O)OCC1CO1	128.13
Acrylate n-butyl	16	CCCCOC(=O)C=C	128.17
Acrylate octyl	17	CCCCCCCCOC(=O)C=C	184.27
Acrylate pentabromophenyl	18	C=CC(=O)OC1=C(C(=C(C(=C1Br)Br)Br)Br)Br	542.60
Acrylic Acid	19	C=CC(=O)O	72.06
Acrylonitrile	20	C=CC#N	53.06
Allyl chloride	21	C=CCCl	76.52
Alpha-Methyl styrene	22	CC(=C)C1=CC=CC=C1	118.18
Anthracene 9-vinyl	23	C=CC1=C2C=CC=CC2=CC3=CC=CC=C31	204.27
Benzophenone, p-vinyl	24	C=CC1=CC=CC=C1C(=O)C2=CC=CC=C2	208.25
Benzyl methacrylate	25	CC(=C)C(=O)OCC1=CC=CC=C1	176.21
Benzylidene cyanoacetate ethyl	26	CCOC(=O)C(=CC1=CC=CC=C1)C#N	201.22
Benzylidene cyanoacetate methyl	27	COC(=O)C(=CC1=CC=CC=C1)C#N	187.19
Benzylidene malononitrile	28	C1=CC=C(C=C1)C=C(C#N)C#N	154.17
Caprolactame N-vinyl	29	C=CN1CCCCCC1=O	139.19
Carbazole N-vinyl	30	C1CCCC3C1C2C(CCCC2)N3C=C	193.24
Cinnamonitrile	31	C1=CC=C(C=C1)C=CC#N	129.16
Dichlorostyrene	32	C=CC1=C(C=CC(=C1)Cl)Cl	173.04
Ethylene trans dichloro	33	C(=CCl)Cl	96.94
Fumaronitrile	34	C(=CC#N)C#N	78.07
Glycidyl methacrylate	35	CC(=C)C(=O)OCC1CO1	142.15
Indene	36	C1C=CC2=CC=CC=C21	116.16
Maleic anhydride	37	C1=CC(=O)OC1=O	98.06
Maleimide	38	C1=CC(=O)NC1=O	97.07
Methacrylamide	39	CC(=C)C(=O)N	85.10
Methacrylate (?-hy-droxymethyl) ethyl	40	CC(CO)OC(=O)C(=C)C	144.17
Methacrylate 2-chloroethyl	41	CC(=C)C(=O)OCCCl	148.59
Methacrylate 2-hydroxyethyl	42	CC(=C)C(=O)OCCO	130.14

Methacrylate benzyl	43	CC(=C)C(=O)OCC1=CC=CC=C1	176.21
Methacrylate butyl	44	CCCCOC(=O)C(=C)C	142.20
Methacrylate decyl	45	CCCCCCCCCCOC(=O)C(=C)C	226.35
Methacrylate ethyl	46	CCOC(=O)C(=C)C	114.12
Methacrylate ethyl ?- benzoyloxy	47	CC(=C)C(=O)OCCOC(=O)C1=CC=CC=C1	234.25
Methacrylate furfuryl	48	CC(=C)C(=O)OCC1=CC=CO1	166.17
Methacrylic acid	49	CC(=C)C(=O)O	86.09
Methacrylonitrile	50	CC(=C)C#N	67.09
Methyl acrylate	51	COC(=O)C=C	86.09
Methyl acrylate ?-(iso-propyl)	52	CC(C)OC(=O)C(=C)C	128.17
Methyl atropate	53	COC(=O)C(=C)C1=CC=CC=C1	162.18
Methyl methacrylate	54	CC(=C)C(=O)OC	100.12
Methyl tiglate	55	CC=C(C)C(=O)OC	114.14
Methyl vinyl ketone	56	CC(=O)C=C	70.09
N-Allyl stearamide	57	C=CCNC(=O)CCCCCCCCCCCCCCCCC	323.56
N-Vinylformamide	58	C=CNC=O	71.08
Phenyl acetylene	59	C#CC1=CC=CC=C1	102.13
Pyrane 2,3-dihydro	60	C1CC=COC1	84.12
Pyridine 2-methyl-5-vinyl	61	CC1=NC=C(C=C1)C=C	119.16
Pyridine 2-vinyl	62	C=CC1=CC=CC=N1	105.14
Pyridine 4-Vinyl	63	C=CC1=CC=NC=C1	105.14
Pyridine 5-ethyl-2-vinyl	64	CCC1=CN=C(C=C1)C=C	133.19
Pyrrolidone N-vinyl	65	C=CN1CCCC1=O	111.14
Sodium acrylate	66	C=CC(=O)[O-].[Na+]	95.05
Styrene	67	C=CC1=CC=CC=C1	104.15
Styrene ?-methyl	68	CC(=C)C1=CC=CC=C1	118.18
Styrene 2,5-dichloro	69	C=CC1=C(C=CC(=C1)Cl)Cl	173.04
Styrene p-iodo	70	C=CC1=CC=C(C=C1)I	230.05
Styrene p-methyl	71	CC1=CC=C(C=C1)C=C	118.18
Succinimide N-vinyl	72	C=CN1C(=O)CCC1=O	125.13
Vinyl acetate	73	CC(=O)OC=C	86.09
Vinyl benzoate	74	C=COC(=O)C1=CC=CC=C1	148.16
Vinyl bromide	75	C=CBr	106.95
Vinyl chloride	76	C=CCl	62.50
Vinyl ethyl ether	77	CCOC=C	72.11
Vinyl isobutyl ether	78	CC(C)COC=C	100.16
Vinyl methyl ketone	79	CC(=O)C=C	70.09
Vinyl n-butyl ether	80	CCCCOC=C	100.16
Vinyl pivalate	81	CC(C)(C)C(=O)OC=C	128.17
Vinylferrocene	82	C=C[C-]1C=CC=C1.[CH-]1[CH-][CH-][CH-][CH-]1.[Fe]	212.07
Vinylidene chloride	83	C=C(Cl)Cl	96.94
Vinylidene cyanide	84	C=C(C#N)C#N	78.07

Table B1: List wih monomer pairs by their unique ID, the number of experiments, the conversion and the r_1 and r_2 values

ID Monomer 1	ID Monomer 2	N. exp	% Conversion	r_1	r_2	Source
5	**12**	7	10	8.0572	0.0467	[133]
5	**14**	6	7.2	5.8971	0.0200	[133]
5	**54**	9	7.5	4.3327	0.0324	[133]
7	**8**	7	-	1.6469	0.1878	[134]
7	**20**	12	-	1.5906	0.5094	[134]
7	**50**	9	-	0.6550	1.1062	[134]
7	**62**	8	-	4.3909	0.0797	[135]
8	**49**	5	26.5	0.6144	0.3351	[136]
10	**35**	7	13	1.2458	0.5915	[137]
11	**35**	5	7.5	0.2975	3.2435	[138]
13	**67**	6	17.1	0.2740	0.9123	[139]
14	**3**	8	38.8	12.8783	0.0001	[140]
14	**42**	5	10.1	0.2942	13.8974	[141]
16	**35**	8	8.1	0.1834	2.7832	[138]
19	**8**	7	10	1.4570	0.6664	[142]
19	**51**	11	-	1.2240	0.8328	[143]
20	**7**	12	-	0.5094	1.5906	[134]
20	**10**	7	6.5	0.9482	1.2677	[144]
20	**12**	5	-	1.0844	0.8309	[145]
20	**12**	5	3.8	1.7401	1.0834	[146]
20	**14**	6	-	1.1948	0.7673	[147]
20	**15**	11	-	1.0693	1.0856	[148]
20	**17**	5	13.2	1.8026	0.8752	[146]
20	**18**	6	7.5	0.8960	1.8290	[144]
20	**19**	8	-	0.4728	2.4237	[149]
20	**21**	5	2.2	3.3727	0.0781	[150]
20	**26**	9	7	75.0199	0.0009	[151]
20	**28**	6	8.5	23.4461	0.0103	[151]
20	**29**	9	7.1	0.2470	0.0197	[152]
20	**31**	7	3.2	6.9270	0.2744	[151]
20	**32**	12	-	0.2185	0.0830	[153]
20	**36**	9	-	0.2617	0.1839	[154]
20	**48**	10	-	0.1840	1.6867	[155]
20	**49**	7	25.7	0.0951	2.5072	[156]
20	**49**	8	-	0.2653	3.4456	[149]
20	**51**	8	21.6	0.4833	0.4530	[157]
20	**51**	5	-	1.2743	0.8707	[145]
20	**51**	5	6.4	1.5521	0.8506	[158]
20	**51**	6	-	0.4812	0.7156	[159]
20	**54**	5	17.7	0.1365	1.3296	[160]
20	**54**	8	6	0.1420	1.1997	[161]
20	**54**	6	-	0.3303	1.4722	[156]
20	**54**	17	11.2	0.2421	1.1430	[162]
20	**56**	6	21.4	0.5860	1.7339	[163]
20	**57**	5	10.5	3.8648	0.1374	[164]
20	**60**	5	7.2	3.8059	0.0076	[165]

20	**61**	9	-	0.1206	0.2948	166
20	**62**		6	0.1081	0.4376	167
20	**63**	7	-	0.1071	0.3439	167
20	**73**	5	1.2	9.3249	0.0001	168
20	**74**	5	5	5.3980	0.0390	150
20	**76**	12	-	1.6600	0.0628	169
20	**76**	23	9	3.8157	0.0476	170
20	**76**	7	6	3.7389	0.0565	171
20	**83**	5	-	0.5698	0.3889	172
20	**83**	18	19	0.8611	0.1657	173
23	**14**	6	1.7	0.2386	3.3312	174
24	**82**	5	5.8	2.0551	0.0001	175
29	**80**	5	77.7	4.1997	0.0010	176
30	**72**	9	5.2	0.3136	1.3394	177
34	**54**	7	8.6	0.0001	10.5615	178
35	**10**	7	13	0.5915	1.2458	136
36	**73**	5	2.4	7.3945	0.1729	179
37	**51**	5	8.9	0.0181	2.9903	180
37	**60**	8	11.8	0.0001	0.0141	165
37	**72**	8	5.3	0.0159	0.1890	177
43	**54**	9	14	1.0383	0.8041	181
43	**67**	10	7.6	0.5136	0.3143	181
50	**54**	5	-	1.0978	0.2650	161
50	**68**	5	4.6	0.3275	0.0921	168
50	**71**	9	8.5	0.2105	0.4080	182
51	**3**	8	46.2	13.7580	0.0001	140
51	**23**	6	1.8	3.0408	0.0940	174
51	**54**	13	8.2	0.3894	2.2563	183
51	**61**	5	-	0.1731	0.9212	184
51	**62**	5	-	0.1749	1.4552	184
51	**63**	6	-	0.2135	1.8457	184
51	**64**	5	-	0.1799	1.2790	166
51	**72**	13	36.5	1.0167	0.3628	185
51	**76**	8	5.9	10.6344	0.0860	171
54	**23**	6	14.7	3.9826	0.0982	174
54	**25**	9	14	0.8041	1.0383	181
54	**29**	8	7.4	5.6505	0.0138	152
54	**37**	8	-	3.5722	0.0001	186
54	**38**	9	14.5	2.8987	0.2036	187
54	**42**	5	7.3	0.2252	0.9617	141
54	**59**	6	25.9	0.8557	0.0084	188
54	**61**	5	-	0.4300	0.5800	166
54	**62**	5	-	0.4250	0.7502	184
54	**63**	5	-	0.5569	0.7697	184
54	**64**	5	-	0.3929	0.6948	184
54	**65**	5	4.6	5.3409	0.0001	189
54	**69**	5	15.7	0.4019	2.3736	190
54	**72**	6	5.7	9.7711	0.0491	191
54	**73**	6	15.1	26.5576	0.0010	147

54	**73**	6	-	29.9331	0.0445	[147]
54	**76**	6	29.6	0.1170	14.2046	[190]
54	**84**	8	8.1	0.0386	0.0262	[192]
58	**8**	6	18.4	0.0366	0.4961	[193]
58	**37**	9	10.3	0.0754	0.0182	[194]
58	**66**	7	18	0.2340	0.5655	[193]
65	**48**	9	3.8	0.0010	4.0713	[195]
65	**77**	5	63.1	0.9102	0.0063	[176]
66	**8**	7	25	0.3772	1.2520	[142]
66	**73**	5	2	14.4800	0.0236	[142]
67	**10**	7	-	0.1398	0.0778	[196]
67	**11**	9	-	0.7269	0.3735	[197]
67	**12**	11	13.5	0.8876	0.1975	[198]
67	**12**	7	9	1.5223	0.3413	[157]
67	**15**	7	-	0.5164	0.1487	[199]
67	**18**	6	-	0.0824	0.1713	[196]
67	**19**	11	-	0.2620	0.0740	[200]
67	**20**	9	-	0.4586	0.0397	[153]
67	**20**	13	14	0.3471	0.0472	[170]
67	**21**	5	2.8	30.1908	0.0010	[201]
67	**23**	8	10.7	2.1537	0.2416	[202]
67	**25**	10	7.6	0.3143	0.5136	[181]
67	**26**	8	5.1	0.4995	0.0001	[203]
67	**27**	5	5.6	0.4954	0.0021	[203]
67	**28**	7	7.3	0.2723	0.0004	[204]
67	**29**	6	1.7	21.2297	0.0259	[152]
67	**30**	8	5.5	7.7866	0.0010	[205]
67	**31**	11	12	2.6667	0.0009	[204]
67	**32**	12	-	0.2571	0.1358	[151]
67	**33**	6	15.2	93.3947	0.4402	[206]
67	**35**	8	10.3	0.5020	0.5657	[207]
67	**38**	10	12.2	0.0658	0.0742	[187]
67	**48**	10	-	0.3354	0.4128	[155]
67	**49**	5	3	0.1374	0.6692	[150]
67	**51**	6	9.8	0.7399	0.1828	[208]
67	**51**	5	10.4	0.7432	0.1774	[158]
67	**54**	7	5	0.4376	0.4873	[209]
67	**54**	6	13.2	0.5200	0.4600	[208]
67	**54**	6	-	0.5844	0.5460	[210]
67	**54**	10	-	0.5238	0.4738	[147]
67	**54**	9	7.7	0.3886	0.4814	[211]
67	**56**	6	7.4	0.2342	0.3451	[212]
67	**59**	5	8.6	0.3428	0.3356	[188]
67	**61**	9	8.7	0.8196	1.0251	[213]
67	**61**	9	-	0.7598	0.8336	[166]
67	**62**	5	-	0.5070	0.8990	[184]
67	**63**	6	-	0.5225	0.7221	[184]
67	**64**	5	-	0.7107	1.0885	[184]
67	**65**	7	0.9	15.4127	0.0001	[189]

67	**65**	11	8.1	6.4422	0.0001	[214]
67	**65**	5	8.2	15.5073	0.0169	[215]
67	**70**	8	-	0.4990	1.0957	[216]
67	**71**	6	5.7	0.8194	0.9496	[217]
67	**72**	7	6.5	11.0780	0.0550	[191]
67	**72**	10	8.1	7.7529	0.1692	[185]
67	**73**	5	-	3.5946	0.0508	[218]
67	**76**	9	4.6	68.9052	0.1722	[170]
67	**83**	20	19	2.1406	0.1249	[170]
67	**84**	8	29	0.0009	0.0017	[192]
68	**48**	9	-	0.2491	0.2861	[155]
70	**67**	6	11.4	0.9588	0.6228	[219]
73	**9**	7	10.5	0.0010	8.5110	[164]
73	**12**	5	-	0.0358	7.5550	[220]
73	**14**	6	-	0.0381	5.1146	[220]
73	**20**	5	1.2	0.0001	9.3249	[168]
73	**21**	6	7.5	0.6600	0.6293	[190]
73	**33**	6	11.9	0.9119	0.0652	[221]
73	**44**	5	-	0.0626	27.4135	[222]
73	**51**	7	-	0.0659	9.4621	[223]
73	**51**	7	-	0.0142	7.0137	[224]
73	**54**	6	-	0.0445	29.9331	[147]
73	**54**	6	-	0.0364	25.6000	[225]
73	**54**	6	15.1	0.0010	26.5576	[192]
73	**57**	8	10.9	0.7818	0.6303	[164]
73	**72**	7	7.4	0.1077	6.3785	[191]
73	**72**	28	47.6	0.2254	66.9242	[185]
73	**75**	5	37.5	0.2175	4.6450	[226]
73	**76**	11	15.5	0.2835	1.9447	[190]
73	**76**	5	-	0.3323	2.5491	[190]
73	**83**	5	6.5	0.0655	5.4385	[227]
73	**84**	6	8.6	0.0024	0.0808	[192]
76	**50**	5	30.4	0.1584	3.8018	[169]
76	**51**	8	5.9	0.0868	10.6793	[228]
76	**75**	6	-	0.8303	1.0696	[229]
76	**78**	6	16	1.9694	0.0010	[150]
76	**83**	10	17	0.0751	2.4337	[230]
76	**83**	12	92	0.1967	6.3559	[170]
76	**84**	10	15.8	0.2859	0.8607	[230]
78	**20**	7	7.3	0.0001	1.0594	[231]
79	**73**	5	-	19.5284	0.2593	[232]
81	**73**	20	36.6	2.1436	0.9148	[233]
82	**24**	5	5.8	0.0001	2.0551	[175]
83	**9**	7	10.9	0.4229	1.3705	[164]
83	**12**	7	7.5	0.9135	0.8583	[146]
83	**17**	6	5.9	0.8377	0.6689	[146]
83	**21**	5	10	2.8579	0.2608	[190]
83	**38**	9	12.2	0.6826	0.4658	[187]
83	**50**	10	-	0.3800	2.5847	[234]

83	**57**	7	8.1	5.9839	0.0001	[164]
83	**72**	8	7.3	1.4838	0.2890	[191]
83	**74**	5	3	5.3509	0.0654	[150]
84	**19**	8	4.6	0.2908	0.2049	[192]
84	**37**	9	16	28.8909	0.0001	[192]
84	**69**	5	10	0.0066	0.0231	[192]
84	**83**	11	19.3	0.0511	0.0111	[192]

Table B2: Monomer pairs with the properties of the first monomer

ID1	ID2	Melt Point 1	Boil Point 1	Refractive Index 1	Heat Capacity 1	Dipole Moment 1	Dielectric Constant 1	Density 1
5	**12**							
5	**14**							
5	**54**							
7	**8**	-87.8	52.3	1.4017	71.3			
7	**20**	-87.8	52.3	1.4017	71.3			
7	**50**	-87.8	52.3	1.4017	71.3			
7	**62**	-87.8	52.3	1.4017	71.3			
8	**49**	85	192.6					1.13
10	**35**							
11	**35**	-90	213.15	1.4332				0.88
13	**67**		183	1.4673				1.0275
14	**3**	-71.2	98.9	1.4068		1.96	6.05	0.9234
16	**23**	-63.6	146.6	1.4185	251		5.25	0.8898
16	**35**	-63.6	146.6	1.4185	251		5.25	0.8898
19	**8**	13.56	142	1.4224	2.022			1.0511
19	**51**	13.56	142	1.4224	2.022			1.0511
20	**7**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**10**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**12**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**12**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**14**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**15**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**17**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**18**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**19**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**21**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**26**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**28**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**29**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**31**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**32**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**36**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**48**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**49**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**49**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**51**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**51**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**51**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**51**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**54**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**54**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**54**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**54**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**56**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**57**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**60**	-83.51	77.2	1.3911		3.92	33	0.8007

20	**61**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**62**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**63**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**73**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**74**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**76**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**76**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**76**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**83**	-83.51	77.2	1.3911		3.92	33	0.8007
20	**83**	-83.51	77.2	1.3911		3.92	33	0.8007
23	**14**							
24	**82**							
29	**80**							
30	**72**	66	154					
14	**42**	-71.2	98.9	1.4068		1.96	6.05	0.9234
34	**54**	96	186	1.4349				0.9416
35	**10**		189	1.448				1.042
36	**73**	-1.45	182.5	1.5768	1.609			0.996
37	**51**	52.56	202		0.396		52.75	0.934
37	**60**	52.56	202		0.396		52.75	0.934
37	**72**	52.56	202		0.396		52.75	0.934
43	**54**							
43	**67**							
50	**54**	-35.8	90	1.4003	1.883	3.69		0.8001
50	**68**	-35.8	90	1.4003	1.883	3.69		0.8001
50	**71**	-35.8	90	1.4003	1.883	3.69		0.8001
51	**3**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
51	**23**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
51	**54**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
51	**61**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
51	**62**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
51	**63**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
51	**64**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
51	**72**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
51	**76**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
54	**23**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**25**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**29**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**37**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**38**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**42**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**59**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**61**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**62**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**63**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**64**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**65**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**69**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**72**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377

54	**73**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**73**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**76**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
54	**84**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
58	**8**	-16	210	1.494				
58	**37**	-16	210	1.494				
58	**66**	-16	210	1.494				
65	**48**	13.5	91.5					1.04
65	**77**	13.5	91.5					1.04
66	**8**	300						1.25
66	**73**	300						1.25
67	**10**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**11**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**12**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**12**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**15**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**18**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**19**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**20**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**20**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**21**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**23**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**25**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**26**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**27**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**28**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**29**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**30**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**31**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**32**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**33**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**35**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**38**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**48**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**49**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**51**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**51**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**54**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**54**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**54**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**54**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**54**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**56**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**59**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**61**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**61**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**62**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**63**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**64**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016

67	**65**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**65**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**65**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**70**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**71**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**72**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**72**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**73**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**76**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**83**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
67	**84**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
68	**48**	-22.36	165.4	1.5386	1.711		2.28	0.9106
70	**67**							
73	**9**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**12**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**14**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**20**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**21**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**33**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**44**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**51**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**51**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**54**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**54**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**54**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**57**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**72**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**72**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**75**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**76**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**76**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**83**	-100.2	72.6	1.3926	1.969	1.79		0.9256
73	**84**	-100.2	72.6	1.3926	1.969	1.79		0.9256
76	**50**	-153.84	-13.8	1.37		1.45		0.9106
76	**51**	-153.84	-13.8	1.37		1.45		0.9106
76	**75**	-153.84	-13.8	1.37		1.45		0.9106
76	**78**	-153.84	-13.8	1.37		1.45		0.9106
76	**83**	-153.84	-13.8	1.37		1.45		0.9106
76	**83**	-153.84	-13.8	1.37		1.45		0.9106
76	**84**	-153.84	-13.8	1.37		1.45		0.9106
78	**20**	-112	83	1.3966			3.34	0.7645
79	**73**	-7	81	1.4081				0.864
81	**73**							
82	**24**	56	82					
83	**9**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
83	**12**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
83	**17**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
83	**21**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
83	**38**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213

83	**50**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
83	**57**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
83	**72**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
83	**74**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
84	**19**							
84	**37**							
84	**69**							
84	**83**							

Table B2: Monomer pairs with the properties of the second monomer

ID1	ID2	Melt Point 2	Boil Point 2	Refractive Index 2	Heat Capacity 2	Dipole Moment 2	Dielectric Constant 2	Density 2
5	**12**	-63.6	146.6	1.4185	251		5.25	0.8898
5	**14**	-71.2	98.9	1.4068		1.96	6.05	0.9234
5	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
7	**8**	85	192.6					1.13
7	**20**	-83.51	77.2	1.3911		3.92	33	0.8007
7	**50**	-35.8	90	1.4003	1.883	3.69		0.8001
7	**62**		159.138	1.5495			9.126	0.9983
8	**49**	14.68	160	1.4314	1.871	1.65		1.0153
10	**35**		189	1.448				1.042
11	**35**		189	1.448				1.042
13	**67**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
14	**3**	51						
16	**23**							
16	**35**		189	1.448				1.042
19	**8**	85	192.6					1.13
19	**51**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
20	**7**	-87.8	52.3	1.4017	71.3			
20	**10**							
20	**12**	-63.6	146.6	1.4185	251		5.25	0.8898
20	**12**	-63.6	146.6	1.4185	251		5.25	0.8898
20	**14**	-71.2	98.9	1.4068		1.96	6.05	0.9234
20	**15**	-41.5	57	1.449				1.1109
20	**17**	-90	227					0.881
20	**18**							
20	**19**	13.56	142	1.4224	2.022			1.0511
20	**21**	-136	44.8	1.4157	1.635	1.94	8.2	0.9376
20	**26**	51		1.5033				1.1076
20	**28**	84						
20	**29**							
20	**31**	22	263.8	1.6013				1.0304
20	**32**	8		1.5798			2.58	1.246
20	**36**	-1.45	182.5	1.5768	1.609			0.996
20	**48**							
20	**49**	14.68	160	1.4314	1.871	1.65		1.0153
20	**49**	14.68	160	1.4314	1.871	1.65		1.0153
20	**51**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
20	**51**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
20	**51**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
20	**51**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
20	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
20	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
20	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
20	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
20	**56**	-7	81.4	1.4081				0.864
20	**57**		456.108	1.458				
20	**60**	-70	85.5	1.4402		1.4		

20	**61**	-14.23	181	1.54				0.978
20	**62**		159.138	1.5495			9.126	0.9983
20	**63**		65	1.5449			10.5	0.98
20	**73**	-100.2	72.6	1.3926	1.969	1.79		0.9256
20	**74**							
20	**76**	-153.84	-13.8	1.37		1.45		0.9106
20	**76**	-153.84	-13.8	1.37		1.45		0.9106
20	**76**	-153.84	-13.8	1.37		1.45		0.9106
20	**83**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
20	**83**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
23	**14**	-71.2	98.9	1.4068		1.96	6.05	0.9234
24	**82**	56	82					
29	**80**	-92	94	1.4026	2.316	1.25		0.7888
30	**72**							
14	**42**	-12	250	1.4515				1.079
34	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
35	**10**							
36	**73**	-100.2	72.6	1.3926	1.969	1.79		0.9256
37	**51**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
37	**60**	-70	85.5	1.4402		1.4		
37	**72**							
43	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
43	**67**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
50	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
50	**68**	-22.36	165.4	1.5386	1.711		2.28	0.9106
50	**71**	-34.1	172.8	1.542				0.9173
51	**3**	51						
51	**23**							
51	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
51	**61**	-14.23	181	1.54				0.978
51	**62**		159.138	1.5495			9.126	0.9983
51	**63**		65	1.5449			10.5	0.98
51	**64**							
51	**72**							
51	**76**	-153.84	-13.8	1.37		1.45		0.9106
54	**23**							
54	**25**							
54	**29**							
54	**37**	52.56	202		0.396		52.75	0.934
54	**38**	94						
54	**42**	-12	250	1.4515				1.079
54	**59**	-45.1	143	1.547		0.656	2.98	0.93
54	**61**	-14.23	181	1.54				0.978
54	**62**		159.138	1.5495			9.126	0.9983
54	**63**		65	1.5449			10.5	0.98
54	**64**							
54	**65**	13.5	91.5					1.04
54	**69**	8		1.5798			2.58	1.246
54	**72**							

54	**73**	-100.2	72.6	1.3926	1.969	1.79		0.9256
54	**73**	-100.2	72.6	1.3926	1.969	1.79		0.9256
54	**76**	-153.84	-13.8	1.37		1.45		0.9106
54	**84**							
58	**8**	85	192.6					1.13
58	**37**	52.56	202		0.396		52.75	0.934
58	**66**	300						1.25
65	**48**							
65	**77**	-115.8	36	1.3767		1.26		0.7589
66	**8**	85	192.6					1.13
66	**73**	-100.2	72.6	1.3926	1.969	1.79		0.9256
67	**10**							
67	**11**	-90	213.15	1.4332				0.88
67	**12**	-63.6	146.6	1.4185	251		5.25	0.8898
67	**12**	-63.6	146.6	1.4185	251		5.25	0.8898
67	**15**	-41.5	57	1.449				1.1109
67	**18**							
67	**19**	13.56	142	1.4224	2.022			1.0511
67	**20**	-83.51	77.2	1.3911		3.92	33	0.8007
67	**20**	-83.51	77.2	1.3911		3.92	33	0.8007
67	**21**	-136	44.8	1.4157	1.635	1.94	8.2	0.9376
67	**23**							
67	**25**							
67	**26**	51		1.5033				1.1076
67	**27**							
67	**28**	84						
67	**29**							
67	**30**	66	154					
67	**31**	22	263.8	1.6013				1.0304
67	**32**	8		1.5798			2.58	1.246
67	**33**	-49.8	47.64	1.4454	1.205		2.14	1.2565
67	**35**		189	1.448				1.042
67	**38**	94						
67	**48**							
67	**49**	14.68	160	1.4314	1.871	1.65		1.0153
67	**51**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
67	**51**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
67	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
67	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
67	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
67	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
67	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
67	**56**	-7	81.4	1.4081				0.864
67	**59**	-45.1	143	1.547		0.656	2.98	0.93
67	**61**	-14.23	181	1.54				0.978
67	**61**	-14.23	181	1.54				0.978
67	**62**		159.138	1.5495			9.126	0.9983
67	**63**		65	1.5449			10.5	0.98
67	**64**							

67	**65**	13.5	91.5					1.04
67	**65**	13.5	91.5					1.04
67	**65**	13.5	91.5					1.04
67	**70**							
67	**71**	-34.1	172.8	1.542				0.9173
67	**72**							
67	**72**							
67	**73**	-100.2	72.6	1.3926	1.969	1.79		0.9256
67	**76**	-153.84	-13.8	1.37		1.45		0.9106
67	**83**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
67	**84**							
68	**48**							
70	**67**	-30.65	145.3	1.544	1.747	0.125	2.4737	0.9016
73	**9**							
73	**12**	-63.6	146.6	1.4185	251		5.25	0.8898
73	**14**	-71.2	98.9	1.4068		1.96	6.05	0.9234
73	**20**	-83.51	77.2	1.3911		3.92	33	0.8007
73	**21**	-136	44.8	1.4157	1.635	1.94	8.2	0.9376
73	**33**	-49.8	47.64	1.4454	1.205		2.14	1.2565
73	**44**	-75	160	1.4242				0.8936
73	**51**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
73	**51**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
73	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
73	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
73	**54**	-47.55	100.6	1.4142	1.91	1.67	6.32	0.9377
73	**57**		456.108	1.458				
73	**72**							
73	**72**							
73	**75**	-139.5	16	1.438	1.007	1.42		1.4933
73	**76**	-153.84	-13.8	1.37		1.45		0.9106
73	**76**	-153.84	-13.8	1.37		1.45		0.9106
73	**83**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
73	**84**							
76	**50**	-35.8	90	1.4003	1.883	3.69		0.8001
76	**51**	-75.6	80.1	1.404	1.845	1.77	7.03	0.9535
76	**75**	-139.5	16	1.438	1.007	1.42		1.4933
76	**78**	-112	83	1.3966			3.34	0.7645
76	**83**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
76	**83**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213
76	**84**							
78	**20**	-83.51	77.2	1.3911		3.92	33	0.8007
79	**73**	-100.2	72.6	1.3926	1.969	1.79		0.9256
81	**73**	-100.2	72.6	1.3926	1.969	1.79		0.9256
82	**24**							
83	**9**							
83	**12**	-63.6	146.6	1.4185	251		5.25	0.8898
83	**17**	-90	227					0.881
83	**21**	-136	44.8	1.4157	1.635	1.94	8.2	0.9376
83	**38**	94						

83	**50**	-35.8	90	1.4003	1.883	3.69		0.8001
83	**57**		456.108	1.458				
83	**72**							
83	**74**							
84	**19**	13.56	142	1.4224	2.022			1.0511
84	**37**	52.56	202		0.396		52.75	0.934
84	**69**	8		1.5798			2.58	1.246
84	**83**	-122.5	31.6	1.4249	1.148	1.34	4.6	1.213

2. Code

```
import sklearn
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error
from sklearn.tree import export_graphviz

import pandas as pd

import numpy as np
from chefboost import Chefboost


df = pd.read_csv('Reactivityratios.csv')
df = pd.get_dummies(df, columns=['SMILES', 'SMILES2'])
df.head()


labels = df['r1']
features= df.drop('r1', axis = 1)
feature_list = list(features.columns)
features = np.nan_to_num(features.astype(np.float32))



train_features, test_features, train_labels, test_labels =
train_test_split(features, labels, test_size = 0.1, random_state = 42, shuffle =
True)



rf = RandomForestRegressor(n_estimators = 75000, criterion = 'mse', random_state
= 42, max_depth = None)

rf.fit(train_features, train_labels);



estimator = rf.estimators_[5]
export_graphviz(estimator, out_file='tree3.dot',
                feature_names = feature_list,
                class_names = labels,
                rounded = True, proportion = False,
                precision = 2, filled = True)
```

```
predictions = rf.predict(test_features)
# Calculate the absolute errors
errors = abs(predictions - test_labels)
# Print out the mean absolute error (mae)
print(errors)
```

Sources

(1) Wouters, O. J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **2020**, *323* (9), 844-853. DOI: 10.1001/jama.2020.1166 (acccessed 1/13/2023).

(2) Savjani, K. T.; Gajjar, A. K.; Savjani, J. K. Drug solubility: importance and enhancement techniques. *ISRN Pharm* **2012**, *2012*, 195727-195727. DOI: 10.5402/2012/195727 PubMed.

(3) Benet, L. Z. The Role of BCS (Biopharmaceutics Classification System) and BDDCS (Biopharmaceutics Drug Disposition Classification System) in Drug Development. *Journal of Pharmaceutical Sciences* **2013**, *102* (1), 34-42. DOI: 10.1002/jps.23359 (acccessed 2019/09/25).

(4) Nikolakakis, I.; Partheniadis, I. Self-Emulsifying Granules and Pellets: Composition and Formation Mechanisms for Instant or Controlled Release. *Pharmaceutics* **2017**, *9* (4), 50. DOI: 10.3390/pharmaceutics9040050.

(5) Vargason, A. M.; Anselmo, A. C.; Mitragotri, S. The evolution of commercial drug delivery technologies. *Nature Biomedical Engineering* **2021**, *5* (9), 951-967. DOI: 10.1038/s41551-021-00698-w.

(6) Rautio, J.; Kumpulainen, H.; Heimbach, T.; Oliyai, R.; Oh, D.; Järvinen, T.; Savolainen, J. Prodrugs: design and clinical applications. *Nature Reviews Drug Discovery* **2008**, *7* (3), 255-270. DOI: 10.1038/nrd2468.

(7) Paddock, C. Brain tumors: Researchers hail soluble aspirin as potential breakthrough. *MedicalNewsToday*, 30 June 2016, 2016. https://www.medicalnewstoday.com/articles/311306.php.

(8) MACDONALD, F. https://www.sciencealert.com/scientists-think-they-ve-developed-a-drug-that-passes-the-blood-brain-barrier. *Sciencealert (Sciencealert.com)*, 2016. https://www.sciencealert.com/scientists-think-they-ve-developed-a-drug-that-passes-the-blood-brain-barrier (accessed 25/09/2019).

(9) Rashid, M.; Malik, M. Y.; Singh, S. K.; Chaturvedi, S.; Gayen, J. R.; Wahajuddin, M. Bioavailability Enhancement of Poorly Soluble Drugs: The Holy Grail in Pharma Industry. (1873-4286 (Electronic)). From 2019.

(10) Marques, E. F.; Silva, B. F. B. Surfactant Self-Assembly. In *Encyclopedia of Colloid and Interface Science*, Tadros, T. Ed.; Springer Berlin Heidelberg, 2013; pp 1202-1241.

(11) Buckinx, A.-L.; Verstraete, K.; Baeten, E.; Tabor, R. F.; Sokolova, A.; Zaquen, N.; Junkers, T. Kinetic Control of Aggregation Shape in Micellar Self-Assembly. *Angewandte Chemie International Edition* **2019**, *0* (0). DOI: 10.1002/anie.201907371 (acccessed 2019/09/04).

(12) Tesauro, D.; Accardo, A.; Diaferia, C.; Milano, V.; Guillon, J.; Ronga, L.; Rossi, F. Peptide-Based Drug-Delivery Systems in Biotechnological Applications: Recent Advances and Perspectives. *Molecules* **2019**, *24* (2), 351. DOI: 10.3390/molecules24020351 PubMed.

(13) Lawatscheck, C.; Pickhardt, M.; Grafl, A.; Linkert, K.; Polster, F.; Mandelkow, E.; Börner, H. G. Gaining Insights into Specific Drug Formulation Additives for Solubilizing a Potential Anti-Alzheimer Disease Drug B4A1. *Macromolecular Bioscience* **2017**, *17* (10), 1700109. DOI: 10.1002/mabi.201700109 (acccessed 2019/09/24).

(14) Maron, E.; Swisher, J. H.; Haven, J.; Meyer, T. Y.; Junkers, T.; Börner, H. G. Learning from peptides to access functional precision polymer sequences. *Angewandte Chemie International Edition* **2019**, *0* (ja). DOI: 10.1002/anie.201902217 (acccessed 2019/06/05).

(15) Kano, T.; Kakinuma, C.; Wada, S.; Morimoto, K.; Ogihara, T. Enhancement of Drug Solubility and Absorption by Copolymers of 2-Methacryloyloxyethyl Phosphorylcholine and n-Butyl Methacrylate. *Drug Metabolism and Pharmacokinetics* **2011**, *26* (1), 79-86. DOI: https://doi.org/10.2133/dmpk.DMPK-10-RG-070.

(16) Widanapathirana, L.; Tale, S.; Reineke, T. M. Dissolution and Solubility Enhancement of the Highly Lipophilic Drug Phenytoin via Interaction with Poly(N-isopropylacrylamide-co-vinylpyrrolidone) Excipients. *Molecular Pharmaceutics* **2015**, *12* (7), 2537-2543. DOI: 10.1021/acs.molpharmaceut.5b00202.

(17) Jenkins, A. D.; Kratochvíl, P.; Stepto, R. F. T.; Suter, U. W. Glossary of basic terms in polymer science (IUPAC Recommendations 1996). **1996**, *68* (12), 2287-2311. DOI: doi:10.1351/pac199668122287 (acccessed 2023-01-30).

(18) Lutz, J.-F. Defining the Field of Sequence-Controlled Polymers. *Macromolecular Rapid Communications* **2017**, *38* (24), 1700582. DOI: 10.1002/marc.201700582 (acccessed 2019/06/11).
(19) Aplan, M. P.; Gomez, E. D. Recent Developments in Chain-Growth Polymerizations of Conjugated Polymers. *Industrial & Engineering Chemistry Research* **2017**, *56* (28), 7888-7901. DOI: 10.1021/acs.iecr.7b01030.
(20) Lutz, J.-F.; Lehn, J.-M.; Meijer, E. W.; Matyjaszewski, K. From precision polymers to complex materials and systems. *Nature Reviews Materials* **2016**, *1*, 16024, Review Article. DOI: 10.1038/natrevmats.2016.24.
(21) Szwarc, M. 'Living' Polymers. *Nature* **1956**, *178* (4543), 1168-1169. DOI: 10.1038/1781168a0.
(22) Rudin, A.; Choi, P. Chapter 9 - Copolymerization. In *The Elements of Polymer Science & Engineering (Third Edition)*, Rudin, A., Choi, P. Eds.; Academic Press, 2013; pp 391-425.
(23) Arrhenius, S. Über die Dissociationswärme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte. *Zeitschrift für Physikalische Chemie* **1889**, *4U* (1), 96-116. DOI: 10.1515/zpch-1889-0408.
(24) Justi, R.; Gilbert, J. K. History and Philosophy of Science through Models: The Case of Chemical Kinetics. *Science Education* **1999**, *8* (3), 287-307. DOI: 10.1023/a:1008645714002.
(25) Marien, Y. W.; Edeleva, M.; Van Steenberge, P. H. M.; D'Hooge, D. R. Chapter Three - Exploiting the pulsed laser polymerization-size exclusion chromatography technique to retrieve kinetic parameters in radical polymerization: State-of-the-art and future challenges. In *Advances in Chemical Engineering*, Moscatelli, D., Sponchioni, M. Eds.; Vol. 56; Academic Press, 2020; pp 59-95.
(26) Mayo, F. R.; Lewis, F. M. Copolymerization. I. A basis for comparing the behavior of monomers in copolymerization; the copolymerization of styrene and methyl methacrylate. *Journal of the American Chemical Society* **1944**, *66* (9), 1594-1601.
(27) Machado, F. Modeling of the Penultimate Unit Effect in Chain-Growth Copolymerizations. *International Journal of Polymer Science* **2019**, *2019*, 2912417. DOI: 10.1155/2019/2912417.
(28) Nicolas, J.; Guillaneuf, Y.; Lefay, C.; Bertin, D.; Gigmes, D.; Charleux, B. Nitroxide-mediated polymerization. *Progress in Polymer Science* **2013**, *38* (1), 63-235. DOI: https://doi.org/10.1016/j.progpolymsci.2012.06.002.
(29) Matyjaszewski, K.; Xia, J. Atom Transfer Radical Polymerization. *Chemical Reviews* **2001**, *101* (9), 2921-2990. DOI: 10.1021/cr940534g.
(30) Perrier, S. 50th Anniversary Perspective: RAFT Polymerization—A User Guide. *Macromolecules* **2017**, *50* (19), 7433-7447. DOI: 10.1021/acs.macromol.7b00767.
(31) BelBruno, J. J. Molecularly Imprinted Polymers. *Chemical Reviews* **2019**, *119* (1), 94-119. DOI: 10.1021/acs.chemrev.8b00171.
(32) Haven, J. J.; Vandenbergh, J.; Kurita, R.; Gruber, J.; Junkers, T. Efficiency assessment of single unit monomer insertion reactions for monomer sequence control: kinetic simulations and experimental observations. *Polymer Chemistry* **2015**, *6* (31), 5752-5765, 10.1039/C5PY00485C. DOI: 10.1039/C5PY00485C.
(33) Clancy, S. Genetic Mutation. *Nature Education* **2008**, *1(1):187*.
(34) Aerts, A.; Lewis, R. W.; Zhou, Y.; Malic, N.; Moad, G.; Postma, A. Light-Induced RAFT Single Unit Monomer Insertion in Aqueous Solution-Toward Sequence-Controlled Polymers. *Macromol Rapid Commun* **2018**, *39* (19), e1800240. DOI: 10.1002/marc.201800240 From NLM.
(35) Koide, Y.; Aoyama, T.; Shimizu, H.; Kitagawa, T.; Miyauchi, R.; Tachibana, H.; Kodaira, T. Development of deep learning chest X-ray model for cardiac dose prediction in left-sided breast cancer radiotherapy. *Scientific Reports* **2022**, *12* (1), 13706. DOI: 10.1038/s41598-022-16583-8.
(36) Ghanzouri, I.; Amal, S.; Ho, V.; Safarnejad, L.; Cabot, J.; Brown-Johnson, C. G.; Leeper, N.; Asch, S.; Shah, N. H.; Ross, E. G. Performance and usability testing of an automated tool for detection of peripheral artery disease using electronic health records. *Scientific Reports* **2022**, *12* (1), 13364. DOI: 10.1038/s41598-022-17180-5.
(37) Xiong, Y.; Ma, Y.; Ruan, L.; Li, D.; Lu, C.; Huang, L.; the National Traditional Chinese Medicine Medical, T. Comparing different machine learning techniques for predicting COVID-19 severity. *Infectious Diseases of Poverty* **2022**, *11* (1), 19. DOI: 10.1186/s40249-022-00946-4.
(38) Turing, A. M. I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind* **1950**, *LIX* (236), 433-460. DOI: 10.1093/mind/LIX.236.433 (acccessed 4/28/2023).

(39) De Donato, L. Deep Learning for Audio Detection and Video Analysis in Railway Applications. 2020.
(40) Copeland, B. J. The modern history of computing. In *Stanford Encyclopedia of Philosophy*, 2008.
(41) Edgar, T. W.; Manz, D. O. Chapter 6 - Machine Learning. In *Research Methods for Cyber Security*, Edgar, T. W., Manz, D. O. Eds.; Syngress, 2017; pp 153-173.
(42) Asimov, I. *I, Robot*; Spectra, 1950.
(43) Sutton, R. S.; Barto, A. G. *Reinforcement learning: An introduction*; The MIT Press, 2017.
(44) Strieth-Kalthoff, F.; Sandfort, F.; Segler, M. H. S.; Glorius, F. Machine learning the ropes: principles, applications and directions in synthetic chemistry. *Chemical Society Reviews* **2020**, *49* (17), 6154-6168, 10.1039/C9CS00786E. DOI: 10.1039/C9CS00786E.
(45) Mardia, K.; Kent, J.; Bibby, J. Multivariate analysis academic press inc. *London) Ltd* **1979**, *15*, 518.
(46) Hastie, T.; Tibshirani, R.; Friedman, J. H.; Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*; Springer, 2009.
(47) Santosa, F.; Symes, W. W. Linear Inversion of Band-Limited Reflection Seismograms. *SIAM Journal on Scientific and Statistical Computing* **1986**, *7* (4), 1307-1330. DOI: 10.1137/0907087 (acccessed 2023/01/30).
(48) Zeng, Z.; Hoshino, Y.; Rodriguez, A.; Yoo, H.; Shea, K. J. Synthetic polymer nanoparticles with antibody-like affinity for a hydrophilic peptide. *ACS Nano* **2010**, *4* (1), 199-204. DOI: 10.1021/nn901256s PubMed.
(49) Tibshirani, R. THE LASSO METHOD FOR VARIABLE SELECTION IN THE COX MODEL. *Statistics in Medicine* **1997**, *16* (4), 385-395, https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3. DOI: https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3 (acccessed 2023/01/30).
(50) Green, S.; Salkind, N. Using SPSS for Windows and Macintosh: analyzing and understanding data. Pearson Prentice Hall. *Upper Saddle River* **2005**.
(51) Williams, M. N.; Grajales, C. A. G.; Kurkiewicz, D. Assumptions of multiple regression: Correcting two misconceptions. *Practical Assessment, Research, and Evaluation* **2013**, *18* (1), 11.
(52) Behler, J. Constructing high-dimensional neural network potentials: A tutorial review. *International Journal of Quantum Chemistry* **2015**, *115* (16), 1032-1050, https://doi.org/10.1002/qua.24890. DOI: https://doi.org/10.1002/qua.24890 (acccessed 2023/01/30).
(53) Baum, Z. J.; Yu, X.; Ayala, P. Y.; Zhao, Y.; Watkins, S. P.; Zhou, Q. Artificial Intelligence in Chemistry: Current Trends and Future Directions. *Journal of Chemical Information and Modeling* **2021**, *61* (7), 3197-3212. DOI: 10.1021/acs.jcim.1c00619.
(54) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-learning-assisted materials discovery using failed experiments. *Nature* **2016**, *533* (7601), 73-76. DOI: 10.1038/nature17439.
(55) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J Chem Theory Comput* **2015**, *11* (5), 2087-2096. DOI: 10.1021/acs.jctc.5b00099.
(56) Coley, C. W.; Thomas, D. A., 3rd; Lummiss, J. A. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **2019**, *365* (6453). DOI: 10.1126/science.aax1566 From NLM.
(57) Kanikkannan, N. Technologies to Improve the Solubility, Dissolution and Bioavailability of Poorly Soluble Drugs. *Journal of Analytical & Pharmaceutical Research* **2018**, *7*. DOI: 10.15406/japlr.2018.07.00198.
(58) Choi, M. J.; Woo, M. R.; Choi, H. G.; Jin, S. G. Effects of Polymers on the Drug Solubility and Dissolution Enhancement of Poorly Water-Soluble Rivaroxaban. *Int J Mol Sci* **2022**, *23* (16). DOI: 10.3390/ijms23169491 From NLM.
(59) Saal, W.; Ross, A.; Wyttenbach, N.; Alsenz, J.; Kuentz, M. Unexpected Solubility Enhancement of Drug Bases in the Presence of a Dimethylaminoethyl Methacrylate Copolymer. *Molecular Pharmaceutics* **2018**, *15* (1), 186-192. DOI: 10.1021/acs.molpharmaceut.7b00804.

(60) Fine-Shamir, N.; Dahan, A. Methacrylate-Copolymer Eudragit EPO as a Solubility-Enabling Excipient for Anionic Drugs: Investigation of Drug Solubility, Intestinal Permeability, and Their Interplay. *Molecular Pharmaceutics* **2019**, *16* (7), 2884-2891. DOI: 10.1021/acs.molpharmaceut.9b00057.
(61) Saal, W. A.-O.; Ross, A.; Wyttenbach, N.; Alsenz, J.; Kuentz, M. A Systematic Study of Molecular Interactions of Anionic Drugs with a Dimethylaminoethyl Methacrylate Copolymer Regarding Solubility Enhancement. (1543-8392 (Electronic)). From 2017 Apr 3.
(62) Karataş, A.; Yüksel, N.; Baykara, T. Improved solubility and dissolution rate of piroxicam using gelucire 44/14 and labrasol. *Il Farmaco* **2005**, *60* (9), 777-782. DOI: https://doi.org/10.1016/j.farmac.2005.04.014.
(63) Al-Hamidi, H.; Obeidat, W. M.; Nokhodchi, A. The dissolution enhancement of piroxicam in its physical mixtures and solid dispersion formulations using gluconolactone and glucosamine hydrochloride as potential carriers. *Pharmaceutical Development and Technology* **2015**, *20* (1), 74-83. DOI: 10.3109/10837450.2013.871029.
(64) Patnaik, S.; Chunduri, L. A. A.; Akilesh, M. S.; Bhagavatham, S. S.; Kamisetti, V. Enhanced dissolution characteristics of piroxicam–Soluplus® nanosuspensions. *Journal of Experimental Nanoscience* **2016**, *11* (12), 916-929. DOI: 10.1080/17458080.2016.1178402.
(65) Ammanage, A.; Rodriques, P.; Kempwade, A.; Hiremath, R. Formulation and evaluation of buccal films of piroxicam co-crystals. *Future Journal of Pharmaceutical Sciences* **2020**, *6* (1), 16. DOI: 10.1186/s43094-020-00033-1.
(66) Cho, C.-F.; Farquhar, C. E.; Fadzen, C. M.; Scott, B.; Zhuang, P.; von Spreckelsen, N.; Loas, A.; Hartrampf, N.; Pentelute, B. L.; Lawler, S. E. A Tumor-Homing Peptide Platform Enhances Drug Solubility, Improves Blood–Brain Barrier Permeability and Targets Glioblastoma. In *Cancers*, 2022; Vol. 14.
(67) Schmied, F. P.; Bernhardt, A.; Moers, C.; Meier, C.; Endres, T.; Klein, S. A Novel Aminomethacrylate-Based Copolymer for Solubility Enhancement-From Radical Polymer Synthesis to Manufacture and Characterization of Amorphous Solid Dispersions. *Polymers (Basel)* **2022**, *14* (7). DOI: 10.3390/polym14071281 From NLM.
(68) Lyons, R. A.; Hutovic, J.; Piton, M. C.; Christie, D. I.; Clay, P. A.; Manders, B. G.; Kable, S. H.; Gilbert, R. G. Pulsed-Laser Polymerization Measurements of the Propagation Rate Coefficient for Butyl Acrylate. *Macromolecules* **1996**, *29* (6), 1918-1927. DOI: 10.1021/ma950747n.
(69) Nikitin, A. N.; Castignolles, P.; Charleux, B.; Vairon, J.-P. Determination of Propagation Rate Coefficient of Acrylates by Pulsed-Laser Polymerization in the Presence of Intramolecular Chain Transfer to Polymer. *Macromolecular Rapid Communications* **2003**, *24* (13), 778-782, https://doi.org/10.1002/marc.200350025. DOI: https://doi.org/10.1002/marc.200350025 (acccessed 2023/05/03).
(70) Barner-Kowollik, C.; Günzler, F.; Junkers, T. Pushing the Limit: Pulsed Laser Polymerization of n-Butyl Acrylate at 500 Hz. *Macromolecules* **2008**, *41* (23), 8971-8973. DOI: 10.1021/ma8020932.
(71) Pirman, T.; Ocepek, M.; Likozar, B. Radical Polymerization of Acrylates, Methacrylates, and Styrene: Biobased Approaches, Mechanism, Kinetics, Secondary Reactions, and Modeling. *Industrial & Engineering Chemistry Research* **2021**, *60* (26), 9347-9367. DOI: 10.1021/acs.iecr.1c01649.
(72) Haehnel, A. P.; Schneider-Baumann, M.; Hiltebrandt, K. U.; Misske, A. M.; Barner-Kowollik, C. Global Trends for kp? Expanding the Frontier of Ester Side Chain Topography in Acrylates and Methacrylates. *Macromolecules* **2013**, *46* (1), 15-28. DOI: 10.1021/ma302319z.
(73) Haehnel, A. P.; Schneider-Baumann, M.; Arens, L.; Misske, A. M.; Fleischhaker, F.; Barner-Kowollik, C. Global Trends for kp? The Influence of Ester Side Chain Topography in Alkyl (Meth)Acrylates − Completing the Data Base. *Macromolecules* **2014**, *47* (10), 3483-3496. DOI: 10.1021/ma500304f.
(74) Ballard, N.; Asua, J. M. Radical polymerization of acrylic monomers: An overview. *Progress in Polymer Science* **2018**, *79*, 40-60. DOI: https://doi.org/10.1016/j.progpolymsci.2017.11.002.
(75) Kockler, K. B.; Haehnel, A. P.; Fleischhaker, F.; Schneider-Baumann, M.; Misske, A. M.; Barner-Kowollik, C. No Apparent Correlation of kp with Steric Hindrance for Branched Acrylates. *Macromolecular Chemistry and Physics* **2015**, *216* (14), 1573-1582,

https://doi.org/10.1002/macp.201500140. DOI: https://doi.org/10.1002/macp.201500140 (acccessed 2022/02/28).
(76) Bordwell, F. G.; Zhang, X.; Alnajjar, M. S. Effects of adjacent acceptors and donors on the stabilities of carbon-centered radicals. *Journal of the American Chemical Society* **1992**, *114* (20), 7623-7629. DOI: 10.1021/ja00046a003.
(77) Cowie, J. M. G.; Arrighi, V. *Polymers: Chemistry and Physics of Modern Materials* Taylor & Francis Group, 2007.
(78) Mavroudakis, E.; Liang, K.; Moscatelli, D.; Hutchinson, R. A. A Combined Computational and Experimental Study on the Free-Radical Copolymerization of Styrene and Hydroxyethyl Acrylate. *Macromolecular Chemistry and Physics* **2012**, *213* (16), 1706-1716, https://doi.org/10.1002/macp.201200165. DOI: https://doi.org/10.1002/macp.201200165 (acccessed 2023/05/03).
(79) Liang, K.; Hutchinson, R. A. The Effect of Hydrogen Bonding on Intramolecular Chain Transfer in Polymerization of Acrylates. *Macromolecular Rapid Communications* **2011**, *32* (14), 1090-1095, https://doi.org/10.1002/marc.201100224. DOI: https://doi.org/10.1002/marc.201100224 (acccessed 2023/05/03).
(80) Patra, T. K. Data-Driven Methods for Accelerating Polymer Design. *ACS Polym Au* **2022**, *2* (1), 8-26. DOI: 10.1021/acspolymersau.1c00035 From NLM.
(81) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28* (1), 235-242. DOI: 10.1093/nar/28.1.235 (acccessed 5/23/2023).
(82) Miccio, L. A.; Schwartz, G. A. From chemical structure to quantitative polymer properties prediction through convolutional neural networks. *Polymer* **2020**, *193*, 122341. DOI: https://doi.org/10.1016/j.polymer.2020.122341.
(83) Webb, M. A.; Jackson, N. E.; Gil, P. S.; de Pablo, J. J. Targeted sequence design within the coarse-grained polymer genome. *Science Advances* **2020**, *6* (43), eabc6216. DOI: doi:10.1126/sciadv.abc6216.
(84) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361* (6400), 360-365. DOI: doi:10.1126/science.aat2663.
(85) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science* **2018**, *4* (2), 268-276. DOI: 10.1021/acscentsci.7b00572.
(86) Huan, T. D.; Mannodi-Kanakkithodi, A.; Ramprasad, R. Accelerated materials property predictions and design using motif-based fingerprints. *Physical Review B* **2015**, *92* (1), 014106. DOI: 10.1103/PhysRevB.92.014106.
(87) Shi, Y.; Yu, M.; Liu, J.; Yan, F.; Luo, Z.-H.; Zhou, Y.-N. Quantitative Structure–Property Relationship Model for Predicting the Propagation Rate Coefficient in Free-Radical Polymerization. *Macromolecules* **2022**, *55* (21), 9397-9410. DOI: 10.1021/acs.macromol.2c01449.
(88) Adams, M. L.; Lavasanifar, A.; Kwon, G. S. Amphiphilic block copolymers for drug delivery. *Journal of Pharmaceutical Sciences* **2003**, *92* (7), 1343-1355. DOI: https://doi.org/10.1002/jps.10397.
(89) Känkänen, V.; Fernandes, M.; Liu, Z.; Seitsonen, J.; Hirvonen, S.-P.; Ruokolainen, J.; Pinto, J. F.; Hirvonen, J.; Balasubramanian, V.; Santos, H. A. Microfluidic preparation and optimization of sorafenib-loaded poly(ethylene glycol-block-caprolactone) nanoparticles for cancer therapy applications. *Journal of Colloid and Interface Science* **2023**, *633*, 383-395. DOI: https://doi.org/10.1016/j.jcis.2022.11.124.
(90) Koch, K. C.; Tew, G. N. Functional antibody delivery: Advances in cellular manipulation. *Advanced Drug Delivery Reviews* **2023**, *192*, 114586. DOI: https://doi.org/10.1016/j.addr.2022.114586.
(91) Dou, J.; Yu, S.; Reddy, O.; Zhang, Y. Novel ABA block copolymers: preparation, temperature sensitivity, and drug release. *RSC Advances* **2023**, *13* (1), 129-139, 10.1039/D2RA05831F. DOI: 10.1039/D2RA05831F.

(92) Fabbri, M.; Guidotti, G.; Soccio, M.; Lotti, N.; Govoni, M.; Giordano, E.; Gazzano, M.; Gamberini, R.; Rimini, B.; Munari, A. Novel biocompatible PBS-based random copolymers containing PEG-like sequences for biomedical applications: From drug delivery to tissue engineering. *Polymer Degradation and Stability* **2018**, *153*, 53-62. DOI: https://doi.org/10.1016/j.polymdegradstab.2018.04.011.
(93) Yorulmaz Avsar, S.; Kyropoulou, M.; Di Leone, S.; Schoenenberger, C.-A.; Meier, W. P.; Palivan, C. G. Biomolecules Turn Self-Assembling Amphiphilic Block Co-polymer Platforms Into Biomimetic Interfaces. *Frontiers in Chemistry* **2019**, *6*, Review.
(94) Lang, C.; Ye, D.; Song, W.; Yao, C.; Tu, Y.-m.; Capparelli, C.; LaNasa, J. A.; Hickner, M. A.; Gomez, E. W.; Gomez, E. D.; et al. Biomimetic Separation of Transport and Matrix Functions in Lamellar Block Copolymer Channel-Based Membranes. *ACS Nano* **2019**, *13* (7), 8292-8302. DOI: 10.1021/acsnano.9b03659.
(95) Salas-Ambrosio, P.; Tronnet, A.; Verhaeghe, P.; Bonduelle, C. Synthetic Polypeptide Polymers as Simplified Analogues of Antimicrobial Peptides. *Biomacromolecules* **2021**, *22* (1), 57-75. DOI: 10.1021/acs.biomac.0c00797.
(96) Gauthier, M. A.; Klok, H.-A. Peptide/protein–polymer conjugates: synthetic strategies and design concepts. *Chemical Communications* **2008**, (23), 2591-2611, 10.1039/B719689J. DOI: 10.1039/B719689J.
(97) Seymour, L. W.; Flanagan, P. A.; Al-Shamkhani, A.; Subr, V.; Ulbrich, K.; Cassidy, J.; Duncan, R. Synthetic Polymers Conjugated to Monoclonal Antibodies: Vehicles for Tumour-Targeted Drug Delivery. *Selective Cancer Therapeutics* **1991**, *7* (2), 59-73. DOI: 10.1089/sct.1991.7.59 (acccessed 2023/02/11).
(98) Lee, J. M.; Koo, M. B.; Lee, S. W.; Lee, H.; Kwon, J.; Shim, Y. H.; Kim, S. Y.; Kim, K. T. High-density information storage in an absolutely defined aperiodic sequence of monodisperse copolyester. *Nature Communications* **2020**, *11* (1), 56. DOI: 10.1038/s41467-019-13952-2.
(99) Kaempf, G.; Loewer, H.; Witman, M. W. Polymers as substrates and media for data storage. *Polymer Engineering & Science* **1987**, *27* (19), 1421-1435, https://doi.org/10.1002/pen.760271902. DOI: https://doi.org/10.1002/pen.760271902 (acccessed 2023/02/11).
(100) Vrijsen, J. H.; Rubens, M.; Junkers, T. Simple and secure data encryption via molecular weight distribution fingerprints. *Polymer Chemistry* **2020**, *11* (40), 6463-6470, 10.1039/D0PY01071E. DOI: 10.1039/D0PY01071E.
(101) Tidwell, P. W.; Mortimer, G. A. An improved method of calculating copolymerization reactivity ratios. *Journal of Polymer Science Part A: General Papers* **1965**, *3* (1), 369-387, https://doi.org/10.1002/pol.1965.100030137. DOI: https://doi.org/10.1002/pol.1965.100030137 (acccessed 2023/05/09).
(102) Alfrey, T., Jr.; Goldfinger, G. The Mechanism of Copolymerization. *The Journal of Chemical Physics* **2004**, *12* (6), 205-209. DOI: 10.1063/1.1723934 (acccessed 5/9/2023).
(103) Fineman, M.; Ross, S. D. Linear method for determining monomer reactivity ratios in copolymerization. *Journal of Polymer Science* **1950**, *5* (2), 259-262, https://doi.org/10.1002/pol.1950.120050210. DOI: https://doi.org/10.1002/pol.1950.120050210 (acccessed 2023/05/10).
(104) Ni, H.; Hunkeler, D. Prediction of copolymer composition drift using artificial neural networks: copolymerization of acrylamide with quaternary ammonium cationic monomers. *Polymer* **1997**, *38* (3), 667-675. DOI: https://doi.org/10.1016/S0032-3861(96)00532-0.
(105) Rintoul, I.; Wandrey, C. Approach To Predict Copolymer Compositions in Case of Variable Monomer Reactivity. *Macromolecules* **2005**, *38* (19), 8108-8115. DOI: 10.1021/ma051010t.
(106) Tan, Z.; Deng, J.; Zhang, S.; Yu, X. Prediction of monomer reactivity in radical copolymerizations from transition state quantum chemical descriptors. *Polímeros* **2013**, *23*.
(107) Shrinivas, K.; Kulkarni, R. P.; Shaikh, S.; Ghorpade, R. V.; Vyas, R.; Tambe, S. S.; Ponrathnam, S.; Kulkarni, B. D. Prediction of Reactivity Ratios in Free Radical Copolymerization from Monomer Resonance–Polarity (Q–e) Parameters: Genetic Programming-Based Models. **2016**, *14* (1), 361-372. DOI: doi:10.1515/ijcre-2014-0039 (acccessed 2023-05-10).
(108) Kazemi, N.; Lessard, B. H.; Marić, M.; Duever, T. A.; Penlidis, A. Reactivity Ratio Estimation in Radical Copolymerization: From Preliminary Estimates to Optimal Design of Experiments. *Industrial & Engineering Chemistry Research* **2014**, *53* (18), 7305-7312. DOI: 10.1021/ie402765k.

(109) Fazakas-Anca, I. S.; Modrea, A.; Vlase, S. Using the Stochastic Gradient Descent Optimization Algorithm on Estimating of Reactivity Ratios. *Materials (Basel)* **2021**, *14* (16). DOI: 10.3390/ma14164764 From NLM.
(110) Johann, T.; Leibig, D.; Grune, E.; Müller, A. H. E.; Frey, H. Effect of the Substituent Position on the Anionic Copolymerization of Styrene Derivatives: Experimental Results and Density Functional Theory Calculations. *Macromolecules* **2019**, *52* (12), 4545-4554. DOI: 10.1021/acs.macromol.9b00747.
(111) Nguyen, T.; Bavarian, M. Machine Learning Approach to Polymerization Reaction Engineering: Determining Monomers Reactivity Ratios. *arXiv preprint arXiv:2301.01231* **2023**.
(112) Van Herck, J.; Junkers, T. Rapid Kinetic Screening via Transient Timesweep Experiments in Continuous Flow Reactors. *Chemistry–Methods* **2022**, *2* (1), e202100090, https://doi.org/10.1002/cmtd.202100090. DOI: https://doi.org/10.1002/cmtd.202100090 (acccessed 2023/02/12).
(113) Schweidtmann, A. M.; Clayton, A. D.; Holmes, N.; Bradford, E.; Bourne, R. A.; Lapkin, A. A. Machine learning meets continuous flow chemistry: Automated optimization towards the Pareto front of multiple objectives. *Chemical Engineering Journal* **2018**, *352*, 277-282. DOI: https://doi.org/10.1016/j.cej.2018.07.031.
(114) Van Herck, J.; Abeysekera, I.; Buckinx, A.-L.; Cai, K.; Hooker, J.; Thakur, K.; Van de Reydt, E.; Voorter, P.-J.; Wyers, D.; Junkers, T. Operator-independent high-throughput polymerization screening based on automated inline NMR and online SEC. *Digital Discovery* **2022**, *1* (4), 519-526, 10.1039/D2DD00035K. DOI: 10.1039/D2DD00035K.
(115) Asua, J. M.; Beuermann, S.; Buback, M.; Castignolles, P.; Charleux, B.; Gilbert, R. G.; Hutchinson, R. A.; Leiza, J. R.; Nikitin, A. N.; Vairon, J.-P.; et al. Critically Evaluated Rate Coefficients for Free-Radical Polymerization, 5. *Macromolecular Chemistry and Physics* **2004**, *205* (16), 2151-2160, https://doi.org/10.1002/macp.200400355. DOI: https://doi.org/10.1002/macp.200400355 (acccessed 2021/10/21).
(116) Kockler, K. B.; Haehnel, A. P.; Junkers, T.; Barner-Kowollik, C. Determining Free-Radical Propagation Rate Coefficients with High-Frequency Lasers: Current Status and Future Perspectives. *Macromol Rapid Commun* **2016**, *37* (2), 123-134. DOI: 10.1002/marc.201500503.
(117) Barner-Kowollik, C.; Beuermann, S.; Buback, M.; Castignolles, P.; Charleux, B.; Coote, M. L.; Hutchinson, R. A.; Junkers, T.; Lacík, I.; Russell, G. T.; et al. Critically evaluated rate coefficients in radical polymerization – 7. Secondary-radical propagation rate coefficients for methyl acrylate in the bulk. *Polymer Chemistry* **2014**, *5* (1), 204-212, 10.1039/C3PY00774J. DOI: 10.1039/C3PY00774J.
(118) Dervaux, B.; Junker, T.; Schneider-Baumann, M.; Du Prez, F. E.; Barner-Kowollik, C. Propagation rate coefficients of isobornyl acrylate, tert-butyl acrylate and 1-ethoxyethyl acrylate: A high frequency PLP-SEC study. *Journal of Polymer Science Part A: Polymer Chemistry* **2009**, *47* (23), 6641-6654, https://doi.org/10.1002/pola.23706. DOI: https://doi.org/10.1002/pola.23706 (acccessed 2022/03/23).
(119) Barner-Kowollik, C.; Bennet, F.; Schneider-Baumann, M.; Voll, D.; Rölle, T.; Fäcke, T.; Weiser, M.-S.; Bruder, F.-K.; Junkers, T. Detailed investigation of the propagation rate of urethane acrylates. *Polymer Chemistry* **2010**, *1* (4), 470-479. DOI: 10.1039/b9py00352e.
(120) Beuermann, S.; Buback, M. Rate coefficients of free-radical polymerization deduced from pulsed laser experiments. *Progress in Polymer Science* **2002**, *27* (2), 191-254. DOI: https://doi.org/10.1016/S0079-6700(01)00049-1.
(121) Beuermann, S.; Buback, M.; Davis, T. P.; Gilbert, R. G.; Hutchinson, R. A.; Kajiwara, A.; Klumperman, B.; Russell, G. T. Critically evaluated rate coefficients for free-radical polymerization, 3. Propagation rate coefficients for alkyl methacrylates. *Macromolecular Chemistry and Physics* **2000**, *201* (12), 1355-1364, https://doi.org/10.1002/1521-3935(20000801)201:12<1355::AID-MACP1355>3.0.CO;2-Q. DOI: https://doi.org/10.1002/1521-3935(20000801)201:12<1355::AID-MACP1355>3.0.CO;2-Q (acccessed 2022/02/28).
(122) Beuermann, S.; Buback, M.; Davis, T. P.; García, N.; Gilbert, R. G.; Hutchinson, R. A.; Kajiwara, A.; Kamachi, M.; Lacík, I.; Russell, G. T. Critically Evaluated Rate Coefficients for Free-Radical Polymerization, 4. *Macromolecular Chemistry and Physics* **2003**, *204* (10), 1338-1350, https://doi.org/10.1002/macp.200390107. DOI: https://doi.org/10.1002/macp.200390107 (acccessed 2022/02/28).

(123) Beuermann, S.; Buback, M.; Davis, T. P.; Gilbert, R. G.; Hutchinson, R. A.; Olaj, O. F.; Russell, G. T.; Schweer, J.; van Herk, A. M. Critically evaluated rate coefficients for free-radical polymerization, 2.. Propagation rate coefficients for methyl methacrylate. *Macromolecular Chemistry and Physics* **1997**, *198* (5), 1545-1560, https://doi.org/10.1002/macp.1997.021980518. DOI: https://doi.org/10.1002/macp.1997.021980518 (acccessed 2022/02/28).
(124) Nitschke, A.; Riemann, L.; Kollenbach, L.; Braun, V.; Buback, M.; Vana, P. Investigation into the Kinetics of n-Pentyl Methacrylate Radical Polymerization. *Macromolecular Chemistry and Physics* **2020**, *221* (1), 1900345, https://doi.org/10.1002/macp.201900345. DOI: https://doi.org/10.1002/macp.201900345 (acccessed 2022/02/28).
(125) Junkers, T.; Koo, S. P. S.; Barner-Kowollik, C. Determination of the propagation rate coefficient of acrylonitrile. *Polymer Chemistry* **2010**, *1* (4), 438-441, 10.1039/C0PY00019A. DOI: 10.1039/C0PY00019A.
(126) Stach, M.; Lacík, I.; Kasák, P.; Chorvát Jr, D.; Saunders, A. J.; Santanakrishnan, S.; Hutchinson, R. A. Free-Radical Propagation Kinetics of N-Vinyl Formamide in Aqueous Solution Studied by PLP–SEC. *Macromolecular Chemistry and Physics* **2010**, *211* (5), 580-593, https://doi.org/10.1002/macp.200900545. DOI: https://doi.org/10.1002/macp.200900545 (acccessed 2022/02/28).
(127) Stach, M.; Lacík, I.; Chorvát, D.; Buback, M.; Hesse, P.; Hutchinson, R. A.; Tang, L. Propagation Rate Coefficient for Radical Polymerization of N-Vinyl Pyrrolidone in Aqueous Solution Obtained by PLP−SEC. *Macromolecules* **2008**, *41* (14), 5174-5185. DOI: 10.1021/ma800354h.
(128) Buback, M.; Gilbert, R. G.; Hutchinson, R. A.; Klumperman, B.; Kuchta, F.-D.; Manders, B. G.; O'Driscoll, K. F.; Russell, G. T.; Schweer, J. Critically evaluated rate coefficients for free-radical polymerization, 1. Propagation rate coefficient for styrene. *Macromolecular Chemistry and Physics* **1995**, *196* (10), 3267-3280, https://doi.org/10.1002/macp.1995.021961016. DOI: https://doi.org/10.1002/macp.1995.021961016 (acccessed 2022/02/28).
(129) Barner-Kowollik, C.; Beuermann, S.; Buback, M.; Hutchinson, R. A.; Junkers, T.; Kattner, H.; Manders, B.; Nikitin, A. N.; Russell, G. T.; van Herk, A. M. Critically Evaluated Rate Coefficients in Radical Polymerization – 8. Propagation Rate Coefficients for Vinyl Acetate in Bulk. *Macromolecular Chemistry and Physics* **2017**, *218* (1), 1600357, https://doi.org/10.1002/macp.201600357. DOI: https://doi.org/10.1002/macp.201600357 (acccessed 2022/02/28).
(130) Hansch, C.; Fujita, T. p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal of the American Chemical Society* **1964**, *86* (8), 1616-1626. DOI: 10.1021/ja01062a035.
(131) Santos-Martins, D.; Forli, S. Charting Hydrogen Bond Anisotropy. *J Chem Theory Comput* **2020**, *16* (4), 2846-2856. DOI: 10.1021/acs.jctc.9b01248 PubMed.
(132) Brocks, J. J.; Beckhaus, H.-D.; Beckwith, A. L. J.; Rüchardt, C. Estimation of Bond Dissociation Energies and Radical Stabilization Energies by ESR Spectroscopy. *The Journal of Organic Chemistry* **1998**, *63* (6), 1935-1943. DOI: 10.1021/jo971940d.
(133) Trumbo, D. L. Copolymerization behavior of 2-vinylbenzofuran: Copolymers of ethyl acrylate, n-butyl acrylate, and methyl methacrylate. *Journal of Polymer Science Part A: Polymer Chemistry* **1991**, *29* (3), 357-360, https://doi.org/10.1002/pola.1991.080290308. DOI: https://doi.org/10.1002/pola.1991.080290308 (acccessed 2023/03/20).
(134) R. C. Schulz, E. K., and W. Kern *Makromol. Chem.* **1962**, *58*, 160-168
(135) S. P. Rao, S. P., S. L. Kapur, and P. K. Iyer. *Polym. Sci. Polym. Lett. Ed.* **1976**, *14*, 513-516
(136) Narasimhaswamy, T.; Sumathi, S.; Reddy, B. s. R. 2,4,6-Tribromophenyl acrylate-co-glycidyl methacrylate polymers: Synthesis, characterization, and reactivity ratios. *Journal of Polymer Science Part A: Polymer Chemistry* **1992**, *30*, 2165-2172. DOI: 10.1002/pola.1992.080301010.
(137) Bakhshi, H.; Jalal, M.; Mehr, Z.; Bouhendi, H.; Kabiri, K. Emulsion Copolymerization of Butyl Acrylate and Glycidyl Methacrylate: Determination of Monomer Reactivity Ratios. *Iranian Polymer Journal (English Edition)* **2010**, *19*.
(138) Dhal, P. K.; Ramakrishna, M. S.; Babu, G. N. Copolymerization of glycidyl methacrylate with alkyl acrylate monomers. *Journal of Polymer Science: Polymer Chemistry Edition* **1982**, *20* (6), 1581-1585, https://doi.org/10.1002/pol.1982.170200618. DOI: https://doi.org/10.1002/pol.1982.170200618 (acccessed 2023/03/20).

(139) Soundararajan, S.; Reddy, B. S. R. Synthesis, characterization and the reactivity ratios of copolymers of cyclohexyl acrylate with styrene and N-vinyl-2-pyrrolidone. *Polymer* **1993**, *34* (10), 2224-2226. DOI: https://doi.org/10.1016/0032-3861(93)90755-Y.
(140) Angelovici, M. M.; Kohn, D. H. Copolymerization of acrylic esters with ethyl α-cyanocinnamate. *Journal of Applied Polymer Science* **1990**, *40* (3-4), 485-494, https://doi.org/10.1002/app.1990.070400315. DOI: https://doi.org/10.1002/app.1990.070400315 (acccessed 2023/03/20).
(141) Litt, M.; Seiner, J. A. The Role of Monomer Charge-Transfer Complexes in Free Radical Copolymerization. IV. Comparison with the Penultimate Hypothesis for Several Simple Systems. *Macromolecules* **1971**, *4* (3), 316-319. DOI: 10.1021/ma60021a012.
(142) Bourdais, J. *Bull. Soc. Chim. Fr.* **1955**, 485-489.
(143) Moţoc, I.; Muşcutariu, I. Monte Carlo Study of Binary Copolymerizations. 1. Ultimate and Penultimate Effect, Applications. *Journal of Macromolecular Science: Part A - Chemistry* **1981**, *15* (1), 75-84. DOI: 10.1080/00222338108066431.
(144) Janovic, Z.; Saric, K.; Voglt, O. Copolymers of Acrylonitrile with Some Brominated Acrylates. *Journal of Macromolecular Science: Part A - Chemistry* **1983**, *19* (8-9), 1137-1152. DOI: 10.1080/00222338308081090.
(145) Iwakura, Y.; Tamikado, T.; Yamaguchi, M.; Takei, K. Copolymerization of acrylonitrile and vinyl pyridine. *Journal of Polymer Science* **1959**, *39* (135), 203-209, https://doi.org/10.1002/pol.1959.1203913516. DOI: https://doi.org/10.1002/pol.1959.1203913516 (acccessed 2023/03/20).
(146) Jordan, E. F.; Doughty, K. M.; Port, W. S. *Journal of Applied Polymer Science* **1960**, *4* (11), 203-206.
(147) Kuo, J.-F.; Chen, C.-Y. A graphic method for the determination of monomer reactivity ratios. *Journal of Applied Polymer Science* **1982**, *27* (7), 2747-2750, https://doi.org/10.1002/app.1982.070270743. DOI: https://doi.org/10.1002/app.1982.070270743 (acccessed 2023/03/20).
(148) Iwakura, Y.; Kurosaki, T.; Nakabayshi, N. *Makromol. Chem.* **1961**, *44/46*, 570-590.
(149) Bajaj, P.; Sen, K.; Bahrami, S. H. Solution polymerization of acrylonitrile with vinyl acids in dimethylformamide. *Journal of Applied Polymer Science* **1996**, *59*, 1539-1550.
(150) Chapin, E. C.; Ham, G. E.; Mills, C. L. Copolymerization. VII. Relative rates of addition of various monomers in copolymerization. *Journal of Polymer Science* **1949**, *4* (5), 597-604, https://doi.org/10.1002/pol.1949.120040505. DOI: https://doi.org/10.1002/pol.1949.120040505 (acccessed 2023/03/20).
(151) Ronel, S. H.; Kohn, D. H. Copolymerization of acrylonitrile. I. Copolymerization with styrene derivatives containing nitrile groups in the side chain. *Journal of Polymer Science Part A-1: Polymer Chemistry* **1969**, *7* (8), 2209-2219, https://doi.org/10.1002/pol.1969.150070817. DOI: https://doi.org/10.1002/pol.1969.150070817 (acccessed 2023/03/20).
(152) N. Cobianu, C. B., B. Marculescu, and I. Amalinei. *Rev. Roum. Chim.* **1973**, *18(2)*, 305-317
(153) Zaldívar, C.; del Sol, O.; Iglesias, G. D. On the preparation of acrylic acid/vinyl acetate copolymers with constant composition — 1. Copolymerization reactivity ratios. *Polymer* **1998**, *39* (1), 245-246. DOI: https://doi.org/10.1016/S0032-3861(97)00363-7.
(154) Tamikado, T. *Makromol. Chem.* **1960**, *38*, 85-95
(155) Goh, S. H.; Lee, S. Y.; Tan, C. T. Y. *Eur. Polym. J.* **1994**, *30*, 489-494.
(156) Simionescu, C.; Asandei, N.; Liga, A. *Makromol. Chem.* **1967**, *110*, 278-290.
(157) Brar, A. S.; Sunita. Determination of microstructure and glass-transition temperature of acrylonitrile-methyl acrylate copolymers by 13C-NMR spectroscopy. *Journal of Polymer Science Part A: Polymer Chemistry* **1992**, *30* (12), 2549-2557, https://doi.org/10.1002/pola.1992.080301209. DOI: https://doi.org/10.1002/pola.1992.080301209 (acccessed 2023/03/20).
(158) Marvel, C. S.; Schwen, R. *Journal of the American Chemical Society* **1957**, *79* (79), 6003-6005.
(159) Joshi, R. M.; Kapur, S. L. *Journal of Polymer Science* **1954**, *14*, 205-510.
(160) Matsuda, M.; Iino, M.; Tokura, N. *Makromol. Chem.* **1963**, *65*, 232-242.
(161) Bixamaiah, B.; Rao, P. R.; Sundaram, E. V. *Journal of Macromolecular Science, Part A* **1992**, *31*, 71-86.

(162) Borchardt, J. K. Calculation of Reactivity Ratios and Sequence Distributions in Copolymers from Monomers 13C-NMR Data. *Journal of Macromolecular Science: Part A - Chemistry* **1985**, *22* (12), 1711-1733. DOI: 10.1080/00222338508063367.
(163) Lewis, F. M.; Walling, C.; Cummings, W.; Briggs, E. R.; Wenisch, W. J. Copolymerization. VII.1 Copolymerizations of Some Further Monomer Pairs. *Journal of the American Chemical Society* **1948**, *70* (4), 1527-1529. DOI: 10.1021/ja01184a069.
(164) Jordan Jr, E. F.; Wrigley, A. N. Monomer reactivity ratios of N-octadecylacrylamide and N-allylstearamide. *Journal of Applied Polymer Science* **1964**, *8* (1), 527-532, https://doi.org/10.1002/app.1964.070080136. DOI: https://doi.org/10.1002/app.1964.070080136 (acccessed 2023/03/20).
(165) Fujimori, K. Copolymerization of 2,3-Dihydropyran and Ethyl Vinyl Ether with Maleic Anhydride. *Journal of Macromolecular Science: Part A - Chemistry* **1975**, *9* (4), 495-504. DOI: 10.1080/00222337508065872.
(166) Mikhael, M. G.; Mokhtar, S. M.; Saad, G. R.; Naoum, M. M.; Elsabee, M. Z. Copolymerization and interaction of acrylonitrile and indene. *Journal of Polymer Science Part A: Polymer Chemistry* **1989**, *27* (1), 185-191, https://doi.org/10.1002/pola.1989.080270116. DOI: https://doi.org/10.1002/pola.1989.080270116 (acccessed 2023/03/20).
(167) Y. Ywakura, T. T., M. Yamaguki, and K. Takei, . *J. Polym. Sci.* **1959**, *39*, 203
(168) Fordyce, R. G.; Chapin, E. C.; Ham, G. E. Copolymerization. V. Relative Monomer Addition Rates in Vinyl Copolymerization. *Journal of the American Chemical Society* **1948**, *70* (7), 2489-2492. DOI: 10.1021/ja01187a053.
(169) Graillat, C.; Guillot, J.; Guyot, A. *J. Macromol. Sci. Chem. A* **1974**, *8* (6), 1099.
(170) Thompson, B. R.; Raines, R. H. *Journal of Polymer Science* **1959**, *41*, 265-274.
(171) Chapin, E. C.; Ham, G. E.; DFordyce, R. G. *Journal of the American Chemical Society* **1948**, *70*, 529-542.
(172) Marker, L.; Sweeting, O. J.; Wepsic, J. C. *Journal of Polymer Science* **1962**, *57*, 855-866.
(173) Camail, M.; Maesano, J. C.; Margaillan, A.; Pautasso, J. P.; Vernet, J. L. Copolymerisation de la N acryloyl (l) alanine avec l'acrylamide. *European Polymer Journal* **1994**, *30* (4), 485-488. DOI: https://doi.org/10.1016/0014-3057(94)90048-5.
(174) Katz, D.; Relis, J. Free-radical copolymerization of 9-vinylanthracene. *Journal of Polymer Science Part A-1: Polymer Chemistry* **1968**, *6* (8), 2079-2083, https://doi.org/10.1002/pol.1968.150060805. DOI: https://doi.org/10.1002/pol.1968.150060805 (acccessed 2023/03/20).
(175) Sanchez, G.; Weill, G.; Knoesel, R. *Makromol. Chem.* **1978**, *179*, 131-146.
(176) Sidel'kovskaya, F. P.; Shostakovskii, M. F.; Ibragimov, F.; Askarov, M. A. The copolymerization of N-vinyl lactams with alkyl vinyl ethers. *Polymer Science U.S.S.R.* **1964**, *6* (9), 1757-1763. DOI: https://doi.org/10.1016/0032-3950(64)90266-7.
(177) Becker, H. H. a. G. *Makromol. Chem.* **1970**, *133*, 1-12
(178) Price, C. C.; Gilbert, R. D. Copolymerization characteristics of fumaronitrile. *Journal of Polymer Science* **1952**, *8* (6), 577-581, https://doi.org/10.1002/pol.1952.120080601. DOI: https://doi.org/10.1002/pol.1952.120080601 (acccessed 2023/03/20).
(179) T. Yala-Lutokanu, G. J., and A. Deluzarche. *Bull. Soc. Chim. Fr.* **1974**, *3-4*, 609-609
(180) De Wilde, M. C.; Smets, G. Copolymérisation de l'Anhydride Maléique avec Différents Monomères Vinyliques. *Journal of Polymer Science* **1950**, *5* (2), 253-258, https://doi.org/10.1002/pol.1950.120050209. DOI: https://doi.org/10.1002/pol.1950.120050209 (acccessed 2023/03/20).
(181) Braun, D.; Disselhoff, G. Kinetics of copolymerizations: 1. Dilatometric investigation of the copolymerizations of benzyl methacrylate, styrene and methyl methacrylate. *Polymer* **1977**, *18* (9), 963-966. DOI: https://doi.org/10.1016/0032-3861(77)90146-X.
(182) Chen, J.; Goh, S. H.; Lee, S. Y.; Siow, K. S. Miscibility behaviour of polymethacrylates with poly(styrene-co-methacrylonitrile). *Polymer* **1994**, *35* (7), 1477-1481. DOI: https://doi.org/10.1016/0032-3861(94)90348-4.
(183) López-González, M. M. C.; Fernández-García, M.; Barrales-Rienda, J. M.; Madruga, E. L.; Arias, C. Sequence distribution and stereoregularity in methyl methacrylate-methyl acrylate

copolymers at high conversions. *Polymer* **1993**, *34* (14), 3123-3128. DOI: https://doi.org/10.1016/0032-3861(93)90648-T.
(184) Tamikado, T. *Journal of Polymer Science* **1960**, *43*, 489-500.
(185) Furukawa, J.; Tsuruta, T.; Yamamoto, N.; Fukutani, H. *Journal of Polymer Science* **1959**, *37*, 215-227.
(186) Blackley, D. C.; Melvill, H. W. *Makromol. Chem.* **1956**, *18/19*, 16-36.
(187) Van Paesschen, G.; Timmerman, D. *Makromol. Chem.* **1964**, *78*, 112-120.
(188) Uzbekova, A. K.; Rozumovskii, V. V. *Vysokomol. Soedin. A* **1972**, *14* (8), 1681-1685.
(189) Braun, D.; Disselhoff, G.; Quella, F. *Makromol. Chem.* **1978**, *179*, 1239-1248.
(190) Agron, P.; Alfrey, T.; Bohrer, J.; Haas, H.; Wechesler, H. *Journal of Polymer Science* **1948**, *3* (2), 157-166.
(191) Hopff, H.; Schlumban, P. *Makromol. Chem.* **1961**, *43*, 173-179.
(192) Gilbert, H.; Miller, F. F.; Averill, S. J.; Carlson, E. J.; Folt, V. L.; Heller, H. J.; Stewart, F. D.; Schmidt, R. F.; Trumbull, H. L. Vinylidene Cyanide. VII. Copolymerization1a. *Journal of the American Chemical Society* **1956**, *78* (8), 1669-1675. DOI: 10.1021/ja01589a049.
(193) Kathmann, E. E. L.; McCormick, C. L. Water-soluble copolymers. 48. Reactivity ratios of N-vinylformamide with acrylamide, sodium acrylate, and n-butyl acrylate. *Macromolecules* **1993**, *26* (19), 5249-5252. DOI: 10.1021/ma00071a041.
(194) Chang, Y.; McCORMICK, C. L. Water-soluble copolymers. 47. Copolymerization of maleic anhydride and N-vinylformamide. *Macromolecules* **1993**, *26*, 4814-4817.
(195) Zaldívar, D.; Peniche, C.; Bulay, A.; Román, J. S. Free radical copolymerization of furfuryl acrylate and 2-hydroxyethyl-methacrylate. *Journal of Polymer Science Part A: Polymer Chemistry* **1993**, *31* (3), 625-631, https://doi.org/10.1002/pola.1993.080310305. DOI: https://doi.org/10.1002/pola.1993.080310305 (acccessed 2023/03/21).
(196) Hagiopol, C. *Copolymerization*; 1999. DOI: https://doi.org/10.1007/978-1-4615-4183-7.
(197) Chatzi, E. G.; Kammona, O.; Kentepozidou, A.; Kiparissides, C. Infrared spectra and compositional analysis of styrene/2-ethylhexyl acrylate copolymers. *Macromolecular Chemistry and Physics* **1997**, *198* (8), 2409-2420, https://doi.org/10.1002/macp.1997.021980805. DOI: https://doi.org/10.1002/macp.1997.021980805 (acccessed 2023/03/21).
(198) Ziaee, F.; Nekoomanesh, M. Monomer reactivity ratios of styrene-butyl acrylate copolymers at low and high conversions. *Polymer* **1998**, *39* (1), 203-207. DOI: https://doi.org/10.1016/S0032-3861(97)00249-8.
(199) Iwakura, Y.; Nakabayshi, N.; Lee, M. H. *Makromol. Chem.* **1967**, *104*, 37-45.
(200) Kerber, R. *Makromol. Chem.* **1966**, *96*, 30-40.
(201) Alfrey, T.; Harrison, J. G. *J. Am. Chem. Soc.* **1946**, *68*, 299-301.
(202) Katz, D. *J. Polym. Sci.* **1963**, *1*.
(203) Gilath, A.; Ronel, S. H.; Shmueli, M.; Kohn, D. H. *Journal of Applied Polymer Science* **1970**, *14*, 1491-1505.
(204) Kreisel, M.; Garbatski, U.; Kohn, D. H. *J. Polym. Sci. A* **1964**, *2*, 105-121.
(205) Fischer, W. H. a. A. *Makromol. Chem.* **1956**, *21* 77-105.
(206) Kudoh, M.; Akutsu, F.; Odagawa, Y.; Naruchi, K.; Miura, M. Polymerizability of methyl α-substituted acrylates containing methoxycarbonyl groups. *Macromolecular Chemistry and Physics* **1994**, *195* (1), 385-390, https://doi.org/10.1002/macp.1994.021950134. DOI: https://doi.org/10.1002/macp.1994.021950134 (acccessed 2023/03/21).
(207) Kocinov, M. F. S. a. I. M. *Plast. Masi.* **1963**, (1), 7-9
(208) Lewis, F. M.; Walling, C.; Cummings, W.; Briggs, E. R.; Mayo, F. R. Copolymerization. IV. Effects of Temperature and Solvents on Monomer Reactivity Ratios. *Journal of the American Chemical Society* **1948**, *70* (4), 1519-1523. DOI: 10.1021/ja01184a066.
(209) Román, J. S.; Madruga, E. L.; Del Puerto, M. A. Microstructure and tacticity in radical copolymerization I. *Die Angewandte Makromolekulare Chemie* **1979**, *78* (1), 129-143, https://doi.org/10.1002/apmc.1979.050780110. DOI: https://doi.org/10.1002/apmc.1979.050780110 (acccessed 2023/03/21).
(210) Madruga, E.; Román, J.; Puerto, M. Radical Copolymerization of Acrylic Monomers. II Effect of Solvent on Radical Copolymerization of Methyl Methacrylate and Styrene. *Journal of Macromolecular Science, Part A* **1979**, *13*, 1105-1115. DOI: 10.1080/00222337908056703.

(211) Ito, K.; Yamashita, Y. Copolymer composition and microstructure. *Journal of Polymer Science Part A: General Papers* **1965**, *3* (6), 2165-2187, https://doi.org/10.1002/pol.1965.100030607. DOI: https://doi.org/10.1002/pol.1965.100030607 (acccessed 2023/03/21).
(212) Tanaka, T. O. a. H. *Polym Sci. Polym Chem Ed.* **1975**, *13(11)*, 2605-2614.
(213) V. G. Ostroverkhof, I. S. V., and V. G. Siniavskii,. *Vysokomol. Soedin. A* **1961**, *3(8),*

, 1197-1203.
(214) Negulescu, I.; Feldman, D.; Simionescu, C. Copolymerization behaviour of some N-vinyl monomers. *Polymer* **1972**, *13* (4), 149-152. DOI: https://doi.org/10.1016/0032-3861(72)90037-7.
(215) Bork, J. F.; Coleman, L. E. Nitrogen-containing monomers. II. Reactivity ratios of n-vinyloxazolidone and N-vinylpyrrolidone with vinyl monomers. *Journal of Polymer Science* **1960**, *43* (142), 413-421, https://doi.org/10.1002/pol.1960.1204314211. DOI: https://doi.org/10.1002/pol.1960.1204314211 (acccessed 2023/03/21).
(216) Braun, V. D.; Ahn, T.-O.; Kern, W. Über die polymerisation von p-jodstyrol. *Die Makromolekulare Chemie* **1962**, *53* (1), 154-172, https://doi.org/10.1002/macp.1962.020530116. DOI: https://doi.org/10.1002/macp.1962.020530116 (acccessed 2023/03/21).
(217) Wiley, R. H.; Davis, B. Tracer techniques for the determination of monomer reactivity ratios. III. Ionization chamber assay and the reactivity ratios for the styrene/p-methylstyrene copolymerization. *Journal of Polymer Science* **1960**, *46* (148), 423-429, https://doi.org/10.1002/pol.1960.1204614811. DOI: https://doi.org/10.1002/pol.1960.1204614811 (acccessed 2023/03/21).
(218) Ramelow, U. S.; Qiu, Q. H. Monomer reactivity ratios in UV-initiated free-radical copolymerization reactions. *Journal of Applied Polymer Science* **1995**, *57* (8), 911-920, https://doi.org/10.1002/app.1995.070570803. DOI: https://doi.org/10.1002/app.1995.070570803 (acccessed 2023/03/21).
(219) D. Braun, D. C., and W. Czerwinski,. *Makromol. Chem.* **1985**, *186*, 1435-1444
(220) Brar, A. S.; Charan, S. Reactivity ratios and microstructure determination of vinyl acetate–alkyl acrylate copolymers by NMR spectroscopy. *Journal of Polymer Science Part A: Polymer Chemistry* **1995**, *33* (1), 109-116, https://doi.org/10.1002/pola.1995.080330113. DOI: https://doi.org/10.1002/pola.1995.080330113 (acccessed 2023/03/21).
(221) Alfrey Jr, T.; Greenberg, S. Experimental study of copolymerization. III. Copolymerization behavior of polychloroethylenes. *Journal of Polymer Science* **1948**, *3* (2), 297-301, https://doi.org/10.1002/pol.1948.120030214. DOI: https://doi.org/10.1002/pol.1948.120030214 (acccessed 2023/03/21).
(222) Brar, A. S.; Charan, S. Reactivity ratios and microstructure of vinyl acetate–butyl methacrylate copolymers: NMR study. *Journal of Applied Polymer Science* **1994**, *51* (4), 669-674, https://doi.org/10.1002/app.1994.070510411. DOI: https://doi.org/10.1002/app.1994.070510411 (acccessed 2023/03/21).
(223) Brar, A. S.; Charan, S. Sequence determination of vinyl acetate–methyl acrylate copolymers by NMR spectroscopy. *Journal of Applied Polymer Science* **1994**, *53* (13), 1813-1822, https://doi.org/10.1002/app.1994.070531311. DOI: https://doi.org/10.1002/app.1994.070531311 (acccessed 2023/03/21).
(224) Noël, L. F. J.; Van Altveer, J. L.; Timmermans, M. D. F.; German, A. L. Determination of the reactivity ratios of methyl acrylate with the vinyl acetate, vinyl 2,2-dimethyl-propanoate, and vinyl 2-ethylhexanoate. *Journal of Polymer Science Part A: Polymer Chemistry* **1994**, *32* (12), 2223-2227, https://doi.org/10.1002/pola.1994.080321202. DOI: https://doi.org/10.1002/pola.1994.080321202 (acccessed 2023/03/21).
(225) Brar, A. S.; Charan, S. Reactivity ratios and microstructure determination of (vinyl acetate)-(methyl methacrylate) copolymers. *European Polymer Journal* **1993**, *29* (5), 755-759. DOI: https://doi.org/10.1016/0014-3057(93)90140-B.
(226) Mayo, F. R.; Walling, C.; Lewis, F. M.; Hulse, W. F. Copolymerization. V.1 Some Copolymerizations of Vinyl Acetate. *Journal of the American Chemical Society* **1948**, *70* (4), 1523-1525. DOI: 10.1021/ja01184a067.
(227) Grassie, N.; McNeill, I. C.; McLaren, I. F. Determination of the reactivity ratios for the copolymerization of vinyl acetate and vinyl chloride using nuclear magnetic resonance spectroscopy.

Journal of Polymer Science Part B: Polymer Letters **1965**, *3* (11), 897-900, https://doi.org/10.1002/pol.1965.110031102. DOI: https://doi.org/10.1002/pol.1965.110031102 (acccessed 2023/03/21).
(228) Chapin, E. C.; Ham, G. E.; Fordyce, R. G. Copolymerization. IV. The Validity of the Tripolymer Equation for the Systems: Styrene—Vinyl Chloride—Methyl Acrylate and Styrene—Vinyl Chloride—Acrylonitrile. *Journal of the American Chemical Society* **1948**, *70* (2), 538-542. DOI: 10.1021/ja01182a032.
(229) Michel, A.; Schmidt, G.; Guyot, A. Vinyl Chloride-Vinyl Bromide Copolymers. *Journal of Macromolecular Science: Part A - Chemistry* **1973**, *7* (6), 1279-1296. DOI: 10.1080/10601327308060498.
(230) Yamashita, Y.; Ito, K.; Ishii, H.; Hoshino, S.; Kai, M. Characterization of Sequence Distribution of Vinylidene Chloride-Vinyl Chloride Copolymers to High Conversions by Nuclear Magnetic Resonance Spectroscopy. *Macromolecules* **1968**, *1* (6), 529-532. DOI: 10.1021/ma60006a015.
(231) Inoue, M.; Otsu, T. Interaction of several polymers with p-substituted phenols in aqueous solution. *Journal of Polymer Science: Polymer Chemistry Edition* **1976**, *14* (8), 1939-1944, https://doi.org/10.1002/pol.1976.170140811. DOI: https://doi.org/10.1002/pol.1976.170140811 (acccessed 2023/03/21).
(232) Inoue, H.; Otsu, T. Radical terpolymerizations of acrylonitrile, vinyl sulfide and vinyl ether. Evaluation of reactivities of vinyl sulfide and vinyl ether toward polyacrylonitrile radical. *Die Makromolekulare Chemie* **1971**, *148* (1), 251-260, https://doi.org/10.1002/macp.1971.021480121. DOI: https://doi.org/10.1002/macp.1971.021480121 (acccessed 2023/03/21).
(233) Haas, H. C.; Simon, M. S. Reactivity ratios of some monomer pairs. *Journal of Polymer Science* **1952**, *9* (4), 309-314, https://doi.org/10.1002/pol.1952.120090403. DOI: https://doi.org/10.1002/pol.1952.120090403 (acccessed 2023/03/21).
(234) Suggate, J. R. Radical copolymerisation of vinylidene chloride and methacrylonitrile, 1. Sequence microstructure measured by 220 MHz 1H NMR. *Die Makromolekulare Chemie* **1978**, *179* (5), 1219-1229, https://doi.org/10.1002/macp.1978.021790510. DOI: https://doi.org/10.1002/macp.1978.021790510 (acccessed 2023/03/21).