

# Machine Learning for the Identification of Hydration Mechanisms of Pharmaceutical-Grade Cellulose Polymers and Their Mixtures with Model Drugs

Przemysław Talik <sup>1,\*</sup> and Aleksander Mendyk <sup>2</sup>

<sup>1</sup> Department of Inorganic and Analytical Chemistry, Faculty of Pharmacy, Jagiellonian University Medical College, 9 Medyczna St., 30-688 Krakow, Poland

<sup>2</sup> Department of Pharmaceutical Technology and Biopharmaceutics, Faculty of Pharmacy, Jagiellonian University Medical College, 9 Medyczna St., 30-688 Krakow, Poland; aleksander.mendyk@uj.edu.pl

\* Correspondence: przemyslaw.talik@uj.edu.pl

**Abstract:** Differently bound water molecules confined in hydrated hydroxypropyl cellulose (HPC) type MF and their mixtures (1:1 *w/w*) with lowly soluble salicylic acid and highly soluble sodium salicylate were investigated by differential scanning calorimetry (DSC). The obtained ice-melting DSC curves of the HPC/H<sub>2</sub>O samples were deconvoluted into multiple components, using a specially developed curve decomposition tool. The ice-melting enthalpies of the individual deconvoluted components were used to estimate the amounts of water in three states in the HPC matrix: free water (FW), freezing bound water (FBW), and non-freezing water (NFW). A search for an optimal number of Gaussian functions was carried out among all available samples of data and was based on the analysis of the minimum fitting error vs. the number of Gaussians. Finally, three Gaussians accounting for three fractions of water were chosen for further analysis. The results of the calculations are discussed in detail and compared to previously obtained experimental DSC data. AI/ML tools assisted in theory elaboration and indirect validation of the hypothetical mechanism of the interaction of water with the HPC polymer.

**Keywords:** machine learning; artificial intelligence; signal decomposition; expectation–maximization algorithm; differential scanning calorimetry; DSC; hydroxypropyl cellulose; non-freezing water; open source

**Citation:** Talik, P.; Mendyk, A. Machine Learning for the Identification of Hydration Mechanisms of Pharmaceutical-Grade Cellulose Polymers and Their Mixtures with Model Drugs. *Appl. Sci.* **2021**, *11*, 7751. <https://doi.org/10.3390/app11167751>

Academic Editor: Luis Javier Garcia Villalba

Received: 8 July 2021

Accepted: 13 August 2021

Published: 23 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

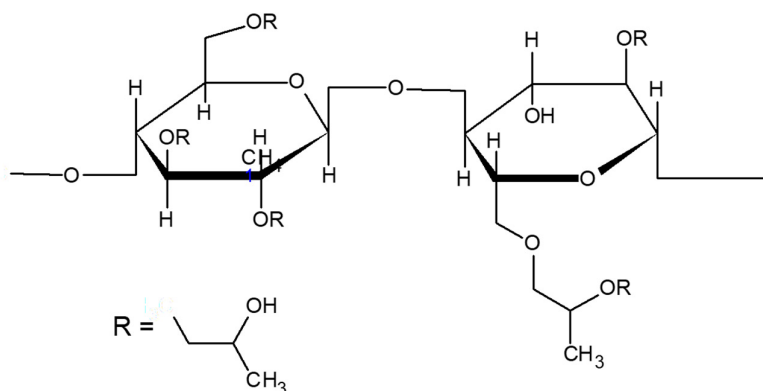
Machine learning (ML) took over the whole artificial intelligence (AI) world, introducing computers' self-adaptation and autonomous learning capabilities to the various areas of science and technology. These applications are not restricted anymore to predictive modeling but are also exploited in the knowledge processing and discovery. For the latter, data mining is a key example of application of ML based on raw data and used for the explanation of possible mechanisms governing observed phenomena. In pharmaceutical industry, where PAT/QbD [1] approaches are now obligatory, this enhancement of data comprehension and explanation is becoming more and more important for rational drug development processes. PAT is a source of increasing amounts of data based on a constantly growing number of available analytical techniques standardized for pharmaceutical applications. One of such techniques is differential scanning calorimetry (DSC). Calorimetry is a fundamental technique for measuring the thermal properties of materials to establish the relationship between temperature and certain physical properties of a compound. Moreover, it is the only method for the direct study of changes in enthalpy of the examined physical processes. Phase transition analysis in modern drug studies has significantly attracted the attention of many research groups dealing with drug discovery and development [2–5]. The study of thermodynamic parameters such as melting ( $T_m$ ),

crystallization ( $T_c$ ), or glass transition temperature ( $T_g$ ), enthalpy ( $\Delta H$ ), and heat capacity ( $C_p$ ) is a source of valuable information that can be used in the development of new drugs and/or in the improvement of those already used in therapy [6–8]. Due to the relative ease of both qualitative and quantitative analysis [9], the DSC technique has its well-established position and priority over others.

AI/ML tools have also found their application in DSC signal analysis, usually in the predictive approach. An example of such application is the work of Wyttenbach et al. [10], where DSC signals were used directly as molecular descriptors for the prediction of solubility of various compounds. Such modeling approach falls into the category of quantitative structure–property relationship (QSPR) and uses DSC results as input variables for a model to predict a physicochemical property of interest—the latter being solubility in this example.

As DSC provides empirical data concerning the thermal behavior of various systems, we still need to understand the possible physical mechanisms governing that behavior. For this purpose, one can use an empirical approach, whereby, based on certain numerical assumptions, the DSC curve is decomposed into simpler elements, and these elements' behavior, namely, parameters of their structure, is a source of knowledge about the whole system. Such signal processing works under the principle of the expectation–maximization algorithm.

Hydroxypropyl cellulose (HPC) (Figure 1) is a water-soluble polymer with a semi-crystalline structure, high degree of amorphous content, and high molecular mobility and plasticity [11]. As other structural derivatives of cellulose, HPC is biocompatible with human tissues [12]. Its physical and chemical properties can be easily modified during synthesis, because the secondary hydroxyl groups in the side chain are available for further etherification with propylene oxides; consequently, the side chains may have more than one propylene substituent. Thus, HPC can be synthesized in various grades, and each grade can correspond to up to six viscosity types. This makes the polymer frequently used in pharmaceutical sciences [13–16] as a drug carrier for oral drug delivery systems.



**Figure 1.** An idealized structure of hydroxypropyl cellulose with a molar substitution of 3.0.

The interaction of water with polysaccharides changes their internal structure and influences their physical, mechanical, and chemical properties. As for other hydrophilic derivatives (cellulose, chitosan, schizophyllan, hyaluronan, carboxymethyl cellulose) [17], the interactions between HPC and water are related to hydrogen bonding and the presence of “nanocavities” formed in the matrix [18]. Depending on water concentration and the grade of the polymer, new stable and reversible structures can form [19]. The existence of three distinct fractions of water is now widely assumed: free (or bulk) water (FW), freezing bound water (FBW), and non-freezing bound water (NFW) [20,21]. Free water does not significantly differ in its melting and crystallization temperature and enthalpy from normal (bulk) water. The second type, freezing bound water, indicates the water

fraction that is less closely associated with the polymer's chains. It reveals a crystallization phase transition and a melting point below 0 °C that distinguishes it from free water. The FBW fraction is also characterized by supercooling and significantly smaller enthalpies than the FW fraction. While first-order phase transitions of small fractions of water that are strongly associated with the polymer chains cannot be observed calorimetrically, we can describe the NFW fraction of water. There are reports showing that this type of water entrapped in a hydrophilic polymer does not crystallize even when cooled down to −100 °C [18]. Moreover, for a low concentration of water, all water found in the matrix is considered non-freezing [18]. The sum of freezing bound and non-freezing bound water fractions is the bound water content.

Two drugs, having distinct properties with respect to solubility, were used as model drugs. Due to structural similarities and their low molecular weight, salicylic acid (SA) and sodium salicylate (NaSA) were selected. Both drugs are often used as model drugs [22–25]. The former is poorly soluble in water, while the latter is freely soluble in water. Aromatic anions, such as salicylate anion, are known to improve the solubility of various types of cellulose [26,27]. Nevertheless, these interesting ionic liquids have not been analyzed.

The aim of this work was to use AI/ML tools for the decomposition of DSC curves into elements representing different fractions of water bound to the HPC polymer. AI/ML tools were used for empirical data analysis, without any physical assumptions, and based on the behavior of the water fractions, some hypotheses elucidating possible mechanisms of interaction between water and cellulose polymers were postulated.

## 2. Materials and Methods

### 2.1. DSC studies

Hydrated hydroxypropyl cellulose (HPC) type MF and its mixtures (1:1 *w/w*) with lowly soluble salicylic acid (SA) or highly soluble sodium salicylate (NaSA) were investigated by differential scanning calorimetry (DSC). The basic physicochemical properties of HPC type MF as well as of salicylic acid and sodium salicylate were previously presented [20]. The same work also presented sample preparation, DSC instrumentation, software, thermal protocol, and other experimental details. The water concentrations of the hydrated samples under study, expressed as  $W_c$ , ranged from 0.2 to 5.0 g/g. The DSC curves obtained in that study, in the section related to melting processes, were the subject of deconvolution analysis. Based on the previously reported results [20], the main principles of the current research were set as follows:

- the water concentrations of the hydrated samples under study were expressed as  $W_c$ , although  $W_c$  is defined as the water fraction  $m_{H_2O}$  related to the dry mass  $m_{dry}$  of raw HPC or an HPC mixture ( $W_c = m_{H_2O}/m_{dry}$ )
- the decomposition of the curves was used for the DSC plots, from the lowest to the highest hydration level  $W_c$
- only melting peaks were used for decomposition analysis
- the amount of NFW water calculated for raw HPC was equal to 0.54 g/g, which is comparable to the amount calculated for HPC/NaSA 0.48 g/g
- NFW calculated for HPC/SA was equal to 0.18 g/g, which is almost 1/3 of the NFW amount measured for raw HPC and the HPC/NaSA mixture
- for water contents below NFW, all water found in the matrix is considered non-freezing NFW
- below the NFW value, the melting peaks for raw HPC and its mixtures are not visible in the DSC measurements; for that reason, the decomposition procedure was performed only for samples with a water concentration  $W_c > \text{NFW}$ , usually, above 0.7–0.8 g/g
- the weight of freezing-bound FBW and non-freezing NFW water per g HPC or HPC/mixture is constant

- the developed model should consider the contribution of specific nanostructures, the so-called nanocavities, in water holding
- the developed model should consider the strong dissociative effect of the Na<sup>+</sup> cation

## 2.2. Curve Decomposition Tool

The basic concept in the approach for building in-house decomposition tool for DSC curves was the simple definition of the target function as a sum of Gaussian functions resulting in a signal compared with the measured one. The error of such comparison was used for the optimization of parameters of the Gaussian functions, namely, center, spread, and height. In our work, non-symmetric or bi-Gaussian functions (NSG) were applied following the Equation (1):

$$NSG = \begin{cases} \text{if } x \leq b; a \cdot \exp\left[-\frac{(x-b)^2}{c_1}\right] \\ \text{if } x > b; a \cdot \exp\left[-\frac{(x-b)^2}{c_2}\right] \end{cases} \quad (1)$$

where

NSG—non-symmetric Gaussian function

$a$ —peak constant

$b$ —center constant

$c_1$  and  $c_2$ —spread constants

The tool was developed as a script of R statistical environment and published on the sourceforge website [28] under GPL v3 license, which means that it is freely available for both personal and commercial use.

The following external packages were used for building R peak decomposer: nloptr [29], GenSA [30], rgenoud [31], and optimx [32]. These packages were optionally chosen for the whole optimization process.

### - nloptr

NLopt is a package for accessing the NLopt system from the level of R script. Out of many optimization algorithms available in NLopt, the general optimization algorithm, the so-called Controlled Random Search (CRS) with local mutation [33], was chosen. It is a first-choice tool for the global optimization loop of the R peak decomposer.

### - GenSA

GenSA stands for “generalized simulated annealing” [34] and introduces another global optimization algorithm based on the simulated annealing approach with fully automated control over starting and operational parameters, like temperature and its decay rate. This is the second-choice tool for the global optimization part of R peak decomposer.

### - rgenoud

Rgenoud is the R Version of GENetic Optimization Using Derivatives. This tool is both a global and a local optimizer using a genetic algorithm for global optimization and subsequently the BFGS method for local fine-tuning of the optimized parameters [35].

### - optimx

The R package optimx is a multi-optimizer using multiple local optimization methods in a simple and elegant interface [36]. This package was implemented with a limited number of optimization methods: BFGS, nlm, Nelder–Mead, and nlminb with follow.on = False setting. These settings impose non-sequential use of the above-mentioned optimization methods within one optimx() run. The final result is chosen as the best achieved solution for the given execution loop, regardless of the actual algorithm being the source of this solution.

The default optimizer for the R peak decomposer is the R optim() function with the BFGS algorithm that is used as the last one in case any optional algorithm was chosen. There is also a global optimization method, SANN—simulated annealing introduced by optim() method, yet there is usually no need to invoke this algorithm due to GenSA’s superior performance in most cases.

The R peak decomposition tool works in the looped manner where the number of iterations in the loop is predetermined together with the stopping criterion—in this case,  $NRMSE = 0.1\%$ , and the loop number was set to 50.

The optional optimizers are used in the cascade mode one after the other, gradually refining the set of Gaussian functions parameters. The user may specify the additional optimization algorithms on his own, yet there is also an automated mode using the following sequence: (1) *nloptr*, (2) *rgenoud*, (3) *gensa*, and (4) *optimx*. This sequence means that all the algorithms are enabled gradually when the tool prediction error is not meeting a stopping criterion. Firstly, the default optimization algorithm BFGS attempts to solve the problem and works for 10% of the number of loops—here, for 5 iterations. If the stopping criterion is not met, then on the iteration no. 6, the *nloptr* takes over and passes its solution to BFGS for a final refinement. If after the next 9 iterations (30% of the maximum iterations number) the stopping criterion is still unmet, then *rgenoud* is enabled, and the system works in the sequence *nloptr*, *rgenoud*, and BFGS. Next, if no improvement is achieved, GenSA is added to the top of the sequence at the iteration no. 25 (50% of the maximum iterations number). Last, *optimx* is added after iteration no. 40 (80% of the maximum iterations number), resulting in the sequence of optimization algorithms: GenSA, *nloptr*, *rgenoud*, *optimx*, and BFGS.

The R peak decomposer reports its results in the form of a text file formatted for an easy import into the spreadsheet. The elements of this report include the raw data of each Gaussian function, parameters of Gaussians, and areas of each Gaussian function.

As the real DSC signal is never ideally positioned in parallel to the temperature axis, a baseline is computed to compensate for this signal distortion. The baseline is expressed in the form of a linear function, whose coefficients may be calculated with two methods: (1) linear regression using signal boundaries at the ends of the temperature axis, (2) non-linear optimization of both coefficients of the linear function carried out jointly with Gaussian functions' coefficients. The latter method was used in this work.

The areas above the curves were computed considering the calculated baselines and used as indicators of the amounts of the respective fractions of water.

In the fitting procedure, the following criteria were kept:

- following Occam's razor, the number of components curves should be minimal
- since the ice-melting curves were asymmetrical, they were deconvoluted using bi-Gaussian functions
- the theoretical curve given by the sum of the individual ones was best fitted to the experimental DSC curve
- the goodness-of-fit criterion was normalized root-mean-squared error ( $NRMSE$ ) calculated according to the formula in Equation (2); an empirical rule of  $NRMSE < 0.1\%$  was employed as an algorithm stopping criterion.

$$NRMSE = \frac{RMSE}{y_{MAX} - y_{MIN}}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i^{PRED} - y_i^{OBS})^2}{n}} \quad (2)$$

where:

$NRMSE$ —normalized root-mean-squared error

$RMSE$ —root-mean-squared error

$y_{iMAX}$ —maximum value of the observed output variable

$y_{iMIN}$ —minimum value of the observed output variable

$y^{PRED}$ —output variable predicted by the model

$y^{OBS}$ —observed output variable

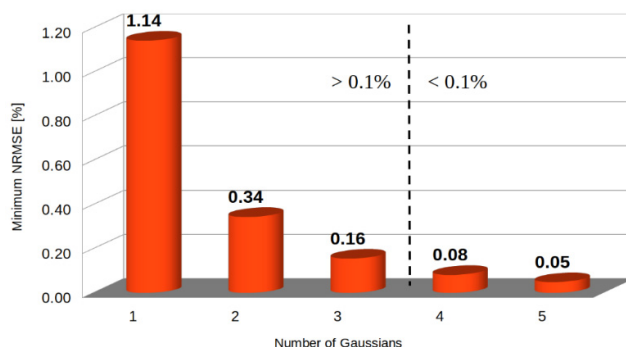
$i$ — $i$ th index of the output variable

$n$ —number of the data points

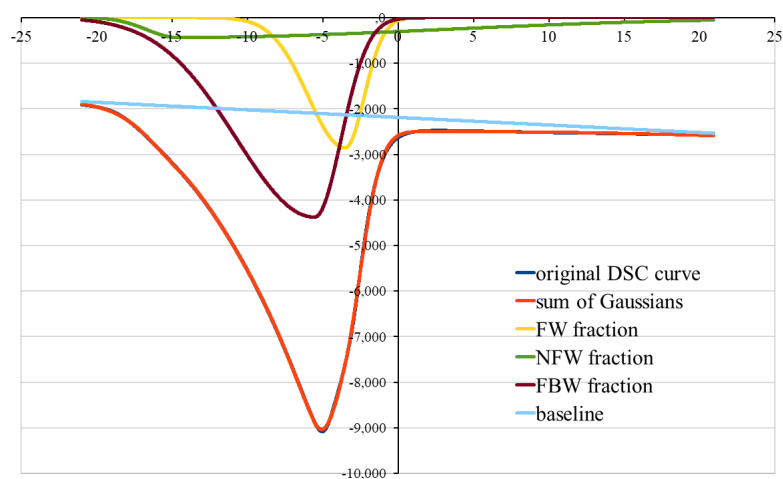
### 3. Results and Discussion

#### 3.1. Search for the Optimum Number of Gaussian Components

The search for an optimal number of Gaussian functions was carried out among all available samples of data and was based on the analysis of the minimum achievable *NRMSE* errors vs. number of Gaussians (Figure 2). In the methodology of the R peak decomposition run, an arbitrary *NRMSE* threshold of 0.1% was applied as a stop criterion for the whole algorithm. In other words, it was decided that 0.1% of *NRMSE* was a threshold of overfitting of a model. Each non-symmetric Gaussian (*NSG*) introduces four parameters, namely, center, peak, and two spread constants (Equation (2)). In Figure 2, a common relationship between prediction error (*NRMSE*) and number of *NSG* is presented—the more the Gaussians, the more accurate the model is. The minimum achievable *NRMSE* approximated 0.1% with three Gaussians and dropped below 0.1% for the four Gaussians; thus, three *NSG* were chosen as the system with best accuracy, yet still preventing overfitting. The optimized number of three *NSG* accounting for three fractions of water was chosen for further analysis. An exemplary decomposition of the DSC curve into three components is shown in Figure 3.



**Figure 2.** Optimal number of Gaussian functions, chosen with the 0.1% *NRSME* criterion.

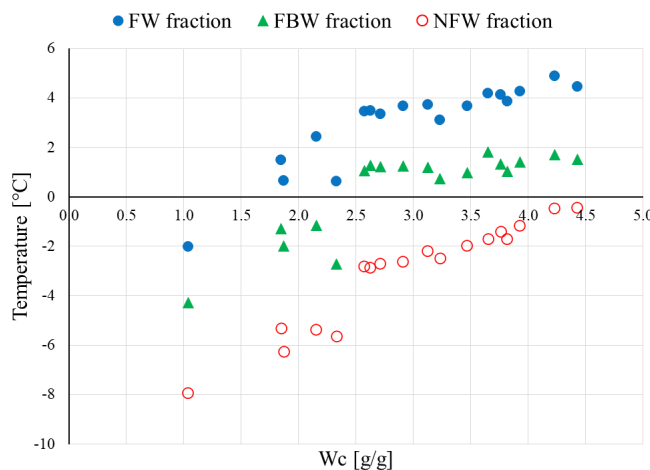


**Figure 3.** Representative three-component decomposition of the DSC curve, obtained from hydrated raw HPC— $W_c = 0.92$  g/g; dark blue line—original DSC curve overlapped by a red line, that stands for the sum of Gaussians; yellow line—FW fraction of water; green line—NFW fraction of water; brown line—FBW fraction of water; light blue line—baseline.

#### 3.2. Three-Component Analysis

As we explained in the introduction, one of the keys to the identification of the appropriate fraction of water is its melting temperature. For this purpose, we assigned each

component a corresponding temperature in this process. Figure 4 shows an example, where determined raw HPC components were grouped in an order determined by the temperature—from the highest to the lowest: first series—the highest temperature—free water fraction FW, second series, fraction of freezing bound water FBW, and third series—last in order, the lowest temperature—NFW fraction. Such assignment allows observing how water fractions are distributed in the polymer matrix or its mixture with drugs.



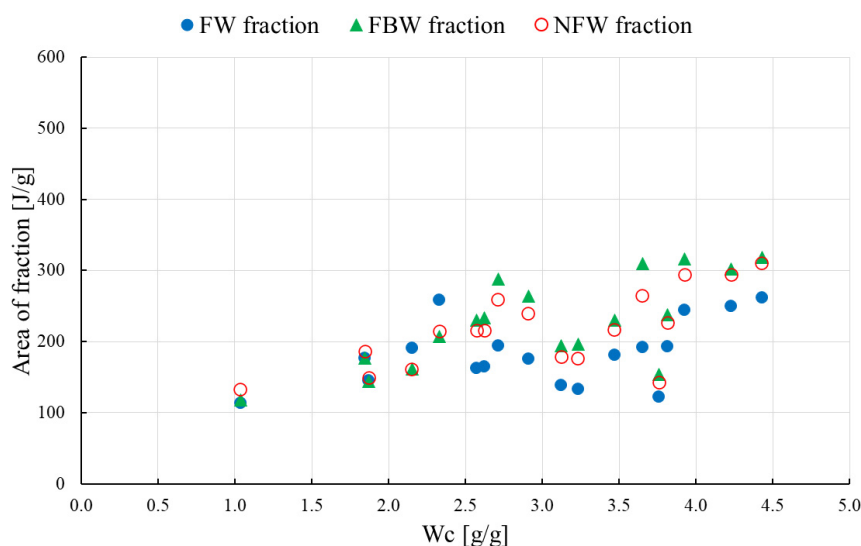
**Figure 4.** Raw HPC—an exemplary graph showing the dependence of the maximum temperature  $T_{\max}$  of FW, FBW, and NFW components on water concentration  $W_c$ .

One can see that the  $T_{\max}$  temperatures of the FW fraction (blue circles) were mostly positive, while the non-freezing water pool (red empty circles) was characterized by negative temperatures. This is due to the fact that the NFW is defined only by water molecules strongly bound by hydrogen bonds and/or embedded in nanocavities of polymer chains. What are nanocavities? FW, FBW, and NFW coexisting in the polymer system form complexes consisting of polymer chains, ice (frozen water), unfrozen water, and air, represented by empty spaces. Both ice and unfrozen water can be trapped in the cavities and/or pores of the matrix. When the dimensions of such a spatial formation do not exceed several Angstroms, the crystallization process is difficult or even impossible. Liu et al. [37] called such hollow spaces “nanocavities” and proved that they are able to form the NFW fraction of water. In this way, they concluded that hydrogen bonding is not the only one factor influencing water crystallization and suggested that it is just one of the possible physical states of hydrated systems. As the NFW pool increases, the interactions between successive layers of water are still strong enough, but they weaken as the water concentration increases. Thus, the temperature  $T_{\max}$  of the water fraction is initially negative and then, when the distances between the binding sites decrease, it becomes positive. In this way, the next fraction of water appears in the matrix. The FBW fraction’s temperatures are negative at low water concentrations (up to around 1.5 g/g), but positive at higher concentrations (>1.5 g/g).

In the next section of this paper, we show the dependence of the surface area of the water fractions, obtained after a deconvolution procedure (grouped using the maximum temperatures  $T_{\max}$ ), on water concentration  $W_c$ .

Figure 5 provides a very interesting observation. The FW fraction (blue circles) is below the other fractions, and for low concentrations—around  $W_c \leq 1.1$  g/g—the concentrations of all water fractions are almost the same. Initially it may seem strange, but such a phenomenon was explained using Raman spectroscopy [38] to study the same samples of hydroxypropyl cellulose type MF. In the conclusions of this work, the authors pointed out that for low water contents equal to 0.69 g/g, all water found in the matrix was considered to be non-freezing. That is why the concentrations of all water fractions were almost the

same. Next, it was mentioned that in the polymer samples with a water concentration of 1.41 g/g, a decrease in FWB water content was found, while the concentrations of the other types of water remained at the same level. This phenomenon can be easily seen in Figure 5. Finally, the authors concluded that for a water content  $W_c \geq 2.2$  g/g, the increase in the concentrations of all types of water began after 16 days. These features seem to be associated with a transition from an anisotropic to an isotropic structure, caused by the clustering of hydrophobic groups.

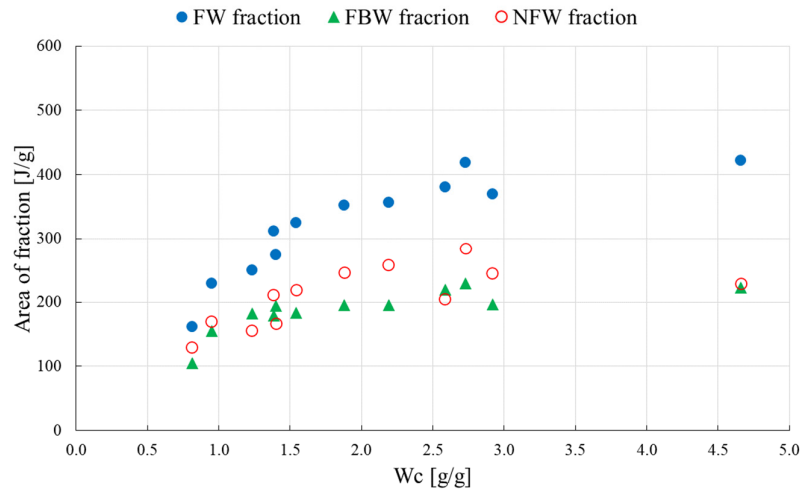


**Figure 5.** The area of different fractions of water, obtained after deconvolution of a raw HPC vs.  $W_c$ . The blue circles indicate the FW fraction, the green triangles the FBW fraction, the red empty circles the NFW fraction.

The above-mentioned process appears in the form of irregular changes in the concentration of each water fraction: an initial increase, then a decrease, and finally a new increase.

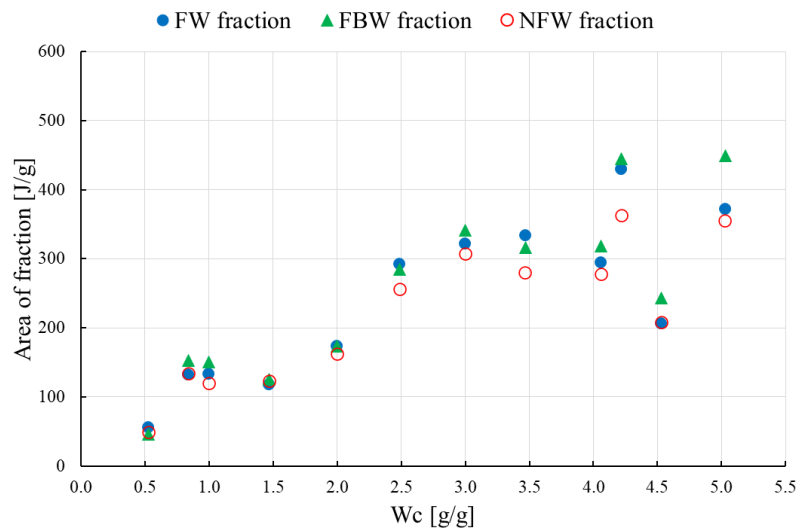
Figure 6 shows the same relationship, Area vs.  $W_c$ , although related to an HPC mixture with lowly soluble salicylic acid SA. Unlike mixtures with sodium salicylate, it is clearly visible how the FBW (green triangles) and NFW (red empty circles) fractions of water became saturated first, while the area associated with free water (blue circles) increased with increasing water concentration in the mixture. This was due to the fact that the content of polymer chains was reduced—HPC + SA 1:1 *w/w*. This lowered the number of possible binding sites and reduced the number of nanocavities. Moreover, some nanocavities could be occupied by small SA molecules. For that reason, the FBW and NFW areas saturated very quickly, and the FW area rapidly increased [21].





**Figure 6.** The area of different fractions of water obtained after deconvolution of HPC/SA mixtures vs.  $W_c$ . The blue circles indicate the FW fraction, the green triangles the FBW fraction, the red empty circles the NFW fraction.

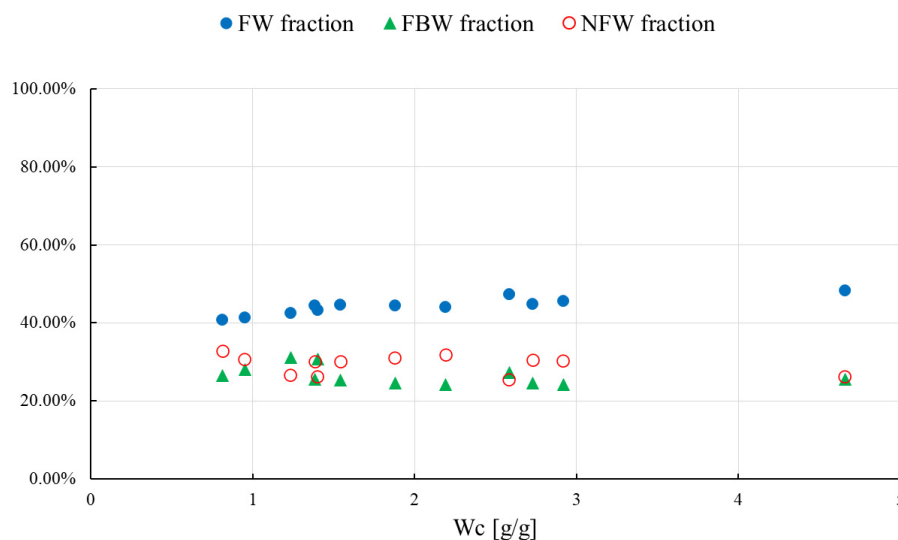
Figure 7 shows a very interesting phenomenon. Similarly to HPC/SA, the number of water binding sites found in HPC/NaSA mixtures was associated with the polymer chains content and was also limited to  $1/2 - \text{HPC} + \text{NaSA} - 1:1$   $w/w$ . However, as it was previously reported [20], the amount of NFW water, calculated for HPC/NaSA, was equal to 0.48 g/g, comparable to the amount of 0.54 g/g calculated for raw HPC. This phenomenon, shown later in Figure 11C, is related to the presence of a strongly dissociating  $\text{Na}^+$  ion [20,37]. The mobility of water molecules in such a system increased so much that it became impossible to distinguish one fraction of water from the other.



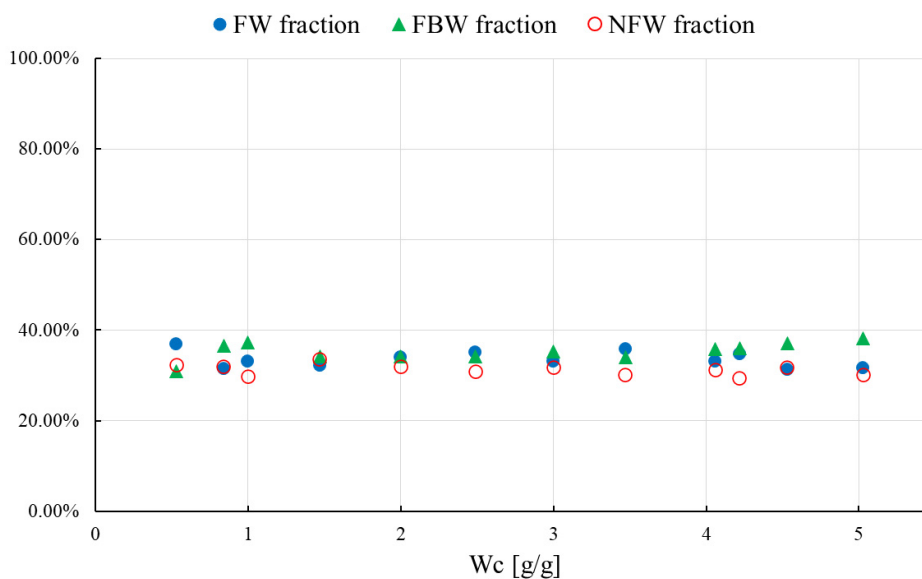
**Figure 7.** The area of different fractions of water, obtained after deconvolution of HPC/NaSA mixtures vs.  $W_c$ . The blue circles indicate the FW fraction, the green triangles the FBW fraction, the red empty circles the NFW fraction.

The relationships between the percentage of the water fractions and  $W_c$  in HPC/SA (Figure 8) and HPC/NaSA mixtures (Figure 9) follows the previously observed relationships between surface area and  $W_c$ . The most important observation is that the percentage

of NFW in the HPC/SA samples was greater than that of the FWB fraction, and the previously observed exchange between water types in HPC/NaSA mixtures was quick and easy.

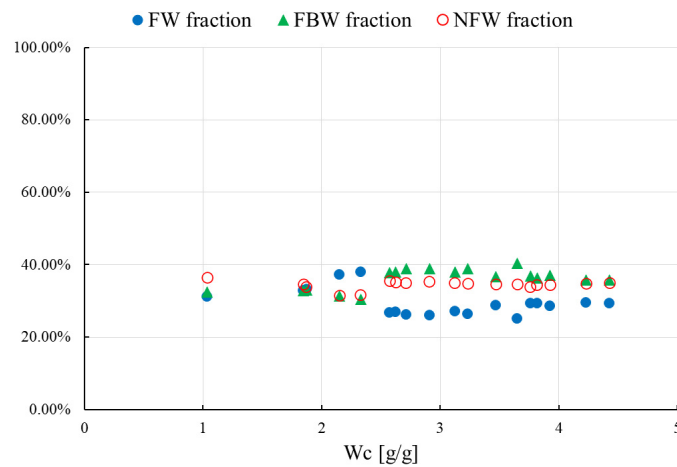


**Figure 8.** HPC/SA mixtures: percentage of the fractions of water in the mixture as a function of water concentration  $W_c$ . The blue circles indicate the FW fraction, the green triangles the FBW fraction, the red empty circles the NFW fraction.



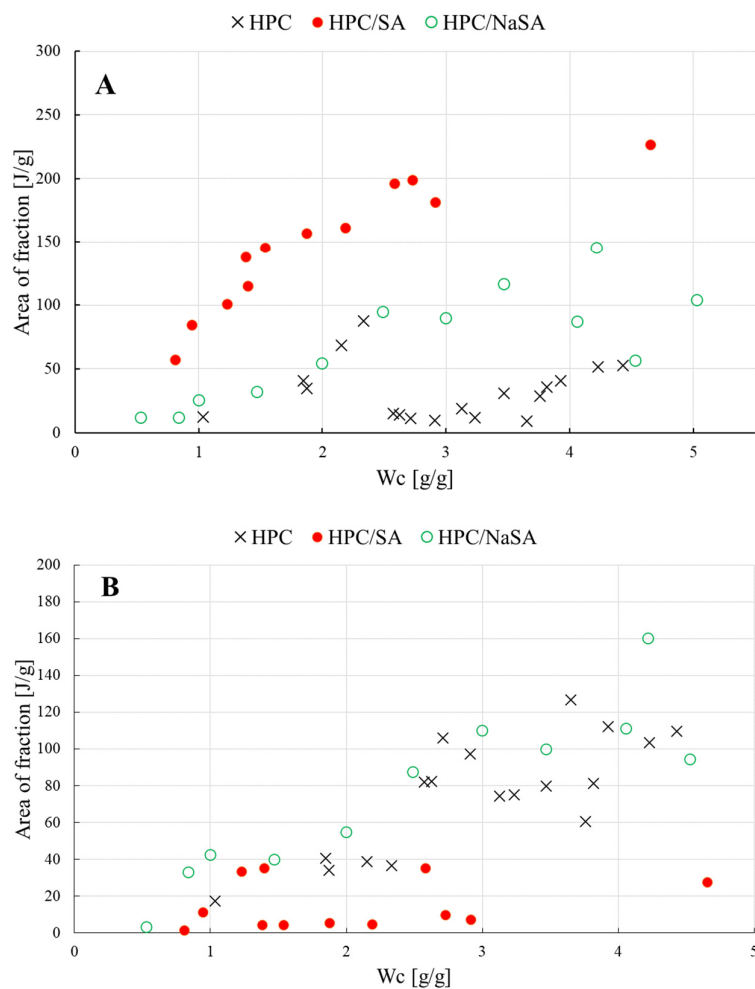
**Figure 9.** HPC/NaSA mixtures: percentage of the fractions of water in the mixture as a function of water concentration  $W_c$ . The blue circles indicate the FW fraction, the green triangles the FBW fraction, the red empty circles the NFW fraction.

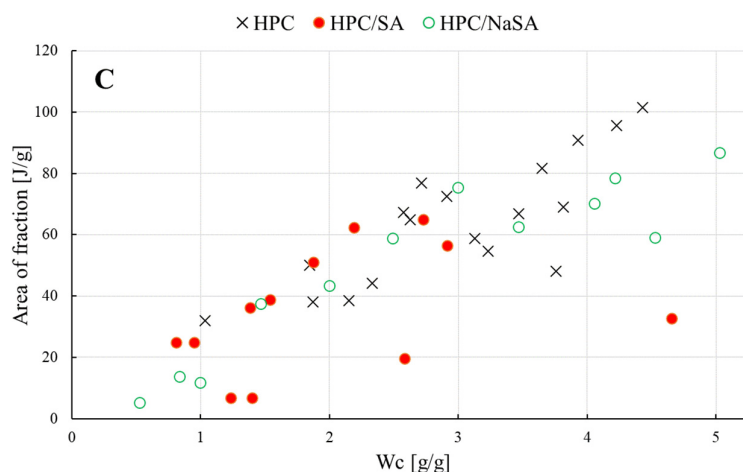
Similarly, it is very thought-provoking that the percentage of FW fraction was the lowest in the raw HPC mixture (Figure 10) when compared to the other NFW and FBW fractions.



**Figure 10.** Raw HPC: percentage of fractions of water in the mixture as a function of water concentration  $W_c$ . The blue circles indicate the FW fraction, the green triangles the FBW fraction, the red empty circles the NFW fraction.

Finally, we compared all FW, FBW, and NFW fractions of water obtained for raw HPC, HPC/SA, and HPC/NaSA mixtures to  $W_c$ , as shown in Figure 11A–C, respectively.





**Figure 11.** Comparison of all fractions (components) of (A) free water (FW), (B) freezing bound water (FBW), (C) non-freezing water (NFW) to  $W_c$ . The red circles, green empty circles, and grey crosses indicate HPC/SA, HPC/NaSA, and raw HPC, respectively.

Such a comparison allowed strengthening the observations made earlier in Figures 5–10. One can see (Figure 11A) that the content of free water FW in raw HPC was the lowest, while in the HPC/SA mixture, it was the highest. Could this be surprising? Certainly not, considering what we wrote about it in the commentary to Figure 6. In a hydrated matrix, where there are few hydrogen binding sites as well as few places capable of trapping water, the FBW and NFW pools saturated faster, while the FW area sharply increased. This corresponded to the content of the FBW fraction, clearly shown in Figure 11B. The strongly dissociating effect of the  $\text{Na}^+$  cation, shown in Figure 11C, was explained while discussing Figure 7.

#### 4. Conclusions

The aim of this work was the use of AI/ML tools for an empirical data analysis of the mutual interactions of water with hydroxypropyl cellulose HPC. AI/ML potential to broaden knowledge was employed to unravel some hidden relationships concerning thermal and ionic phenomena. We were able to confirm the assisting role of AI/ML in the formulation of hypotheses and their at least partial verification and/or falsification in relation to the physical phenomena observed with commonly applicable analytical methods—in our case, a DSC assay.

The main conclusions of the paper are:

- assigning the maximum temperatures  $T_{\max}$  to the empirically determined components allowed confirming the identity of each water fraction—the same conclusions regarding the relationship  $T_{\max}$  vs.  $W_c$ , but obtained from raw, not decomposed DSC curves, were the subject of another publication [20]
- it was shown that the location of the components determined during AI/ML analysis was justified by the Raman spectrometry method [36,37]
- as a result of the method used, the effects of drugs with different solubilities were very well differentiated and shown. The location of the components as well as quantitative aspects were confirmed in the DSC experiments carried out earlier [20,21,38]
- a significant contribution of nanocavities in the formation of strongly bound non-freezing water (NFW) was indicated—the obtained results were confirmed in previous publications [18,20,21,29,37,39]
- it was pointed out the influence of strongly dissociating  $\text{Na}^+$  ions on some quantitative relationships associated with strongly bound non-freezing water (NFW) [20,40]

The presented three-component analysis met our all assumptions and expectations. The obtained theoretical model confirmed the physical properties of hydrated HPC—both raw and in a mixture with drugs of different solubility—previously obtained using differential scanning calorimetry DSC. Such results provide a very promising and proven instrument for modelling hydrated polysaccharide systems, also mixed with small-molecule drugs with varying solubility. The obtained results encourage to perform a decomposition analysis using the same algorithm on polymers with a different structure and physicochemical properties, also in mixtures with macromolecular drugs or proteins.

This type of research may contribute to a better understanding of the release of both highly and poorly soluble drugs from tablet formulations and their dissolution kinetics. Moreover, based on the well-known biopharmaceutical paradigm of drugs fate in the body LADME, the identification of the driving forces and mechanisms of drug release is crucial to the understanding and prediction of drugs' bioavailability. The latter is a primary endpoint for many types of clinical trials and plays a pivotal role in the assessment of drugs' efficacy and safety—an obligatory assessment required for every drug introduced to the market.

The major achievements and novelty of this work include:

- the development and publication of the AI/ML tool for the decomposition of DSC curves as an open source software freely available both for personal and for commercial use [28]
- the proof-of-concept use of the above-mentioned tool for DSC results for HPC/water mixtures with model drugs
- the confirmation of some AI/ML empirically developed hypotheses in the literature about the thermal behavior of HPC/water mixtures

**Author Contributions:** Conceptualization, P.T. and A.M.; methodology, P.T. and A.M.; software, A.M.; formal analysis, P.T.; investigation, A.M.; data curation, P.T. and A.M.; writing—original draft preparation, P.T. and A.M.; writing—review and editing, P.T.; visualization, P.T. and A.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Software: R peak decomposer and Database (database.zip). Available online: <https://sourceforge.net/projects/r-peak-decomposer> (accessed on 1 July 2021).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

PAT	process analytical technologies
QbD	quality by design
DSC	differential scanning calorimetry
HPC	hydroxypropyl cellulose
FW	free (or bulk) water
FBW	freezing bound water
NFW	non-freezing bound water
SA	salicylic acid
NaSA	sodium salicylate
HPC/SA	a binary physical mixture of hydroxypropyl cellulose and salicylic acid 1:1 <i>w/w</i>
HPC/NaSA	a binary physical mixture of hydroxypropyl cellulose and sodium salicylate 1:1 <i>w/w</i>
BFGS	Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm is an iterative method for solving unconstrained nonlinear optimization problems

LADME      Liberation, Absorption, Distribution, Metabolism, Excretion—major processes responsible for drugs behavior in the human body.

## References

1. ICH. The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, Quality Guideline Q8(R2) Pharmaceutical Development. 2009. Available online: [http://www.ich.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Quality/Q8\\_R1/Step4/Q8\\_R2\\_Guideline.pdf](http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf) (accessed on 21 June 2021).
2. Clas, S.D.; Dalton, C.R.; Hancock, B.C. Differential scanning calorimetry: Applications in drug development. *Pharm. Sci. Technol. Today* **1999**, *2*, 311–320, doi:10.1016/S1461-534700181-9.
3. Witika, B.A.; Aucamp, M.; Mweetwa, L.L.; Makoni, P.A. Application of Fundamental Techniques for Physicochemical Characterizations to Understand Post-Formulation Performance of Pharmaceutical Nanocrystalline Materials. *Crystals* **2021**, *11*, 310, doi:10.3390/cryst11030310.
4. Knopp, M.M.; Löbmann, K.; Elder, D.P.; Rades, T.; Holm, R. Recent advances and potential applications of modulated differential scanning calorimetry (mDSC) in drug development. *Eur. J. Pharm. Sci.* **2016**, *87*, 164–173, doi:10.1016/j.ejps.2015.12.024.
5. Leyva-Porras, C.; Cruz-Alcantar, P.; Espinosa-Solis, V.; Martínez-Guerra, E.; Piñón-Balderrama, C.I.; Compean Martínez, I.; Saavedra-Leos, M.Z. Application of Differential Scanning Calorimetry (DSC) and Modulated Differential Scanning Calorimetry (MDSC) in Food and Drug Industries. *Polymers* **2020**, *12*, 5, doi:10.3390/polym12010005.
6. Bernardes, C.E.S.; Joseph, A.; Minas da Piedade, M.E. Some practical aspects of heat capacity determination by differential scanning calorimetry. *Thermochim. Acta* **2020**, *687*, 178574, doi:10.1016/j.tca.2020.178574.
7. Farooq, D. Theoretical Aspects of Differential Scanning Calorimetry as a Tool for the Studies of Equilibrium Thermodynamics in Pharmaceutical Solid Phase Transitions. *AAPS PharmSciTech* **2016**, *17*, 572–577, doi:10.1208/s12249-016-0530-2.
8. Aloisio, C.; Gomes De Oliveira, A.; Longhi, M. Characterization, inclusion mode, phase-solubility and in vitro release studies of inclusion binary complexes with cyclodextrins and meglumine using sulfamerazine as model drug. *Drug Dev. Ind. Pharm.* **2014**, *40*, 919–928, doi:10.3109/03639045.2013.790408.
9. Talik, P.; Żuromska-Witek, B.; Hubicka, U.; Krzek, J. The use of the DSC method in quantification of active pharmaceutical ingredients in commercially available one component tablets. *Acta Pol. Pharm.* **2017**, *74*, 1049–1055.
10. Wyttenbach, N.; Niederquell, A.; Kuentz, M. Machine Estimation of Drug Melting Properties and Influence on Solubility Prediction. *Mol. Pharm.* **2020**, *17*, 2660–2671, doi:10.1021/acs.molpharmaceut.0c00355.
11. Picker-Freyer, K.M.; Durig, T. Physical mechanical and tablet formation properties of hydroxypropylcellulose: In pure form and in mixtures. *AAPS PharmSciTech* **2007**, *8*, 82–90.
12. Arca, H.C.; Mosquera-Giraldo, L.I.; Bi, V.; Xu, D.; Taylor, L.S.; Edgar, K.J. Pharmaceutical applications of cellulose ethers and cellulose ether esters. *Biomacromolecules* **2018**, *19*, 2351–2376.
13. Klose, R.E.; Glickman, M.; Gums. In *Handbook of Food Additives*, 2nd ed.; Furia, T.E., CRC Press: Boca Raton, FL, USA, 1973; Volume 1, pp. 295–360.
14. Tanaka, A.; Furubayashi, T.; Tomisaki, M.; Kawakami, M.; Kimura, S.; Inoue, D.; Kusamori, K.; Katsumi, H.; Sakane, T.; Yamamoto, A. Nasal drug absorption from powder formulations: The effect of three types of hydroxypropyl cellulose (HPC). *Eur. J. Pharm. Sci.* **2017**, *96*, 284–289.
15. Macchi, E.; Zema, L.; Pandey, P.; Gazzaniga, A.; Felton, L.A. Influence of temperature and relative humidity conditions on the pan coating of hydroxypropyl cellulose molded capsules. *Eur. J. Pharm. Biopharm.* **2016**, *100*, 47–57.
16. Hughey, J.R.; Keen, J.M.; Bennett, R.C.; Obara, S.; McGinity, J.W. The incorporation of low-substituted hydroxypropyl cellulose into solid dispersion systems. *Drug Dev. Ind. Pharm.* **2015**, *41*, 1294–1301.
17. Mičoch, T.; Kučerík, J. Hydration and drying of various polysaccharides studied using DSC. *J. Therm. Anal. Calorim.* **2013**, *113*, 1177–1185.
18. Ping, Z.H.; Nguyen, Q.T.; Chen, S.M.; Zhou, J.Q.; Ding, Y.D. States of water in different hydrophilic polymers—DSC and FTIR studies. *Polymer* **2001**, *42*, 8461–8467.
19. Hatakeyama, T.; Inui, Y.; Iijima, M.; Hatakeyama, H. Bound water restrained by nanocellulose fibres. *J. Therm. Anal. Calorim.* **2013**, *113*, 1019–1025.
20. Talik, P.; Hubicka, U. The DSC approach to study non-freezing water contents of hydrated hydroxypropylcellulose (HPC): A study over effects of viscosity and drug addition. *J. Therm. Anal. Calorim.* **2018**, *132*, 445–451.
21. Talik, P.; Hubicka, U. A study of the drying behaviour of various types of hydrated hydroxypropyl cellulose (HPC) and their mixtures with drugs of different solubility using DSC. *J. Therm. Anal. Calorim.* **2021**, *143*, 247–254.
22. Singh, B.; Chauhan, G.S.; Sharma, D.K.; Kant, A.; Gupta, I.; Chauhan, N. The release dynamics of model drugs from the psyllium and N-hydroxymethylacrylamide based hydrogels. *Int. J. Pharm.* **2006**, *325*, 15–25, doi:10.1016/j.ijpharm.2006.06.007.
23. Pade, V.; Stavchansky, S. Estimation of the Relative Contribution of the Transcellular and Paracellular Pathway to the Transport of Passively Absorbed Drugs in the Caco-2 Cell Culture Model. *Pharm. Res.* **1997**, *14*, 1210–1215, doi:10.1023/A:1012111008617.
24. Tungprapa, S.; Jangchud, I.; Supaphol, P. Release characteristics of four model drugs from drug-loaded electrospun cellulose acetate fiber mats. *Polymer* **2007**, *48*, 5030–5041, doi:10.1016/j.polymer.2007.06.061.
25. Taepaiboon, P.; Rungsardthong, U.; Supaphol, P. Drug-loaded electrospun mats of poly(vinyl alcohol) fibres and their release characteristics of four model drugs. *Nanotechnology* **2006**, *17*, 2317–2329. DOI 10.1088/0957-4484/17/9/041.

26. Brehm, M.; Pulst, M.; Kressler, J.; Sebastiani, D. Triazolium-Based Ionic Liquids: A Novel Class of Cellulose Solvents. *J. Phys. Chem. B* **2019**, *123*, 3994–4003, doi:10.1021/acs.jpcc.8b12082.
27. Brehm, M.; Radicke, J.; Pulst, M.; Shaabani, F.; Sebastiani, D.; Kressler, J. Dissolving Cellulose in 1,2,3-Triazolium- and Imidazolium-Based Ionic Liquids with Aromatic Anions. *Molecules* **2020**, *25*, 3539, doi:10.3390/molecules25153539.
28. R Peak Decomposer. Available online: <https://sourceforge.net/projects/r-peak-decomposer> (accessed on 21 June 2021).
29. R Interface to NLOpt. Available online: <https://cran.r-project.org/package=nloptr> (accessed on 21 June 2021).
30. Generalized Simulated Annealing. Available online: <https://cran.r-project.org/package=GenSA> (accessed on 21 June 2021).
31. R Version of GENetic Optimization Using Derivatives. Available online: <https://cran.r-project.org/package=rngenoud> (accessed on 21 June 2021).
32. R Optimx Package. Available online: <https://cran.r-project.org/package=optimx> (accessed on 21 June 2021).
33. Kaelo, P.; Ali, M.M. Some variants of the controlled random search algorithm for global optimization. *J. Optim. Theory Appl.* **2006**, *130*, 253–264.
34. Xiang, Y.; Gubian, S.; Suomela, B.; Hoeng, J. Generalized Simulated Annealing for Efficient Global Optimization: The GenSA Package for R. *R J.* **2013**, *5*, 13–28. Available online: <https://journal.r-project.org/archive/2013/RJ-2013-002/index.html> (accessed on 21 June 2021).
35. Walter, R.; Mebane, J.; Sekhon, J.S. Genetic Optimization Using Derivatives: The rngenoud Package for R. *J. Stat. Softw.* **2011**, *42*, 1–26. Available online: <http://www.jstatsoft.org/v42/i11/> (accessed on 21 June 2021).
36. Nash, J.C.; Varadhan, R. Unifying Optimization Algorithms to Aid Software System Users: Optimx for R. *J. Stat. Softw.* **2011**, *43*, 1–14. Available online: <https://www.jstatsoft.org/v43/i09/> (accessed on 21 June 2021).
37. Liu, G.L.; Yao, K.D. What causes the unfrozen water in polymers: Hydrogen bonds between water and polymer chains? *Polymer* **2001**, *42*, 3943–3947.
38. Talik, P.; Moskal, P.; Proniewicz, L.M.; Weselucha-Birczyńska, A. The Raman spectroscopy approach to the study of Water–Polymer interactions in hydrated hydroxypropyl cellulose (HPC). *J. Mol. Struct.* **2020**, *1210*, 128062.
39. Park, S.; Venditti, R.A.; Jameel, H.; Pawlak, J.J. Studies of the heat of vaporization of water associated with cellulose fibres characterized by thermal analysis. *Cellulose* **2007**, *14*, 195–204.
40. Berthold, J.; Desbrieres, J.; Rinaudo, M.; Salmén, L. Types of adsorbed water in relation to the ionic groups and their counterions for some cellulose derivatives. *Polymer* **1994**, *35*, 5729–5736.