Polymer Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Polym. Chem.*, 2024, **15**, 1273

Predicting the solubility of gases, vapors, and supercritical fluids in amorphous polymers from electron density using convolutional neural networks†

Oleg I. Gromov

A twin convolutional neural network is proposed to predict the pressure and temperature-dependent sorption of gases, vapors, and supercritical fluids in amorphous polymers, using spatial electron density distribution. These distributions are obtained as 3D tensors (images) from DFT calculations. The dataset, compiled from over 250 literature sources, comprises nearly 15 000 experimental measurements of 79 gases' uptakes (0.01–50 wt%) in 102 different polymers. These measurements, spanning pressures from 1×10^{-3} to 7×10^{2} bar and temperatures from 233 to 508 K, include nearly 500 solvent–polymer systems, ranging from low-pressure sorption in membrane glassy polymers to high-pressure solubility of supercritical fluids in molten polymers. The irreducible mean absolute percentage error (MAPE) is estimated to be around 20%, with a brief discussion on the sources of data variability. In 150 epochs, the model achieved a 32% MAPE on a test set of 1600 measurements concerning 22 polymers not previously encountered by the model.

Received 9th September 2023,
Accepted 17th February 2024

DOI: 10.1039/d3py01028g

rsc.li/polymers

1. Introduction

The ability of polymers to absorb gases, vapors, and supercritical fluids (SCFs) is used or considered in numerous practical applications. Here are a few examples:

(1) **Gas separation and purification**: Polymers are frequently used as membrane materials. The solubility ratio of gases in these polymers is one of the key parameters determining their separation efficiency.[1]

(2) **Sensor technology**: In the development of gas sensors, the solubility of gases in polymers can significantly influence the sensitivity and selectivity of the sensor materials.[2]

(3) **Packaging industry**: Understanding gas solubility in polymers plays a crucial role in food packaging. It controls the permeation of gases like oxygen and carbon dioxide, which in turn affects food freshness and shelf life.[3]

(4) **SCF-assisted polymer processing**: The dissolution of supercritical fluids in polymers enables various processes. This includes foaming polymers, facilitating extrusion, and acting as solvents for doping polymers.[4–7]

An enormous amount of data on solubility in polymers is constantly being accumulated.[8–11] Rationalizing experimental results and predicting the solubility of gases and SCFs are crucial for adding value to the available data, aiding in the screening of existing materials, and designing new ones. Most existing models are parametric, varying in their derivation approximations, applicability scope, and predictive capabilities. A comprehensive review of these parametric models is available elsewhere.[12] The advantages of such models include ease of implementation, a limited number of clearly interpretable parameters, and the ability to utilize tabulated parameters from extensive literature data. However, these benefits come with drawbacks. Many popular models, like the Dual-Mode Sorption model[13–17] and the Flory–Huggins model,[18,19] are domain-specific, applicable either to glassy (thermodynamically non-equilibrium) or to rubbery/liquid (thermodynamically equilibrium) polymer states. Despite numerous specific extensions, a lack of generalization persists, with some notable exceptions like the Non-Equilibrium Thermodynamics for Glassy Polymers (NET-GP) approach.[20–22] For instance, the Non-Equilibrium Lattice Fluid (NELF) model, an extension of the Sanchez–Lacombe (SL) Equation of State (EoS).[23] Secondly, due to their approximate nature, the parameters of these models, although meaningful, are not directly linked to measurable or computable polymer properties. Typically derived from fitting experimental data, it is common to find multiple parameter sets for the same system in the literature. This ambiguity significantly impedes attempts to estimate model parameters from first prin-

Chemistry Department, Lomonosov Moscow State University, Leninskiye Gory 1-3, Moscow, 119991, Russia. E-mail: aalchm@gmail.com

†Electronic supplementary information (ESI) available: Collection of experimentally measured solvent uptakes (CSV), list of polymers (CSV), list of solvents (CSV), optimized geometries (XYZ), polymers' repeating units and solvents in SMILES format (SMI). See DOI: https://doi.org/10.1039/d3py01028g

ciples and limits their predictive capabilities to systems already studied experimentally.[24] An interesting exception is the COSMO-SAC model,[25] based on the COSMO-RS method,[26] which allows the direct application of theoretical calculations but is primarily intended for homogeneous solutions.

At the opposite end of the physical realism scale, molecular modeling methods are widely used to calculate various macroscopic properties, including the solubility of gases in polymers.[12,27] These methods encompass various forms of Molecular Dynamics (MD) and Monte Carlo (MC) simulations. Molecular modeling is becoming both more accurate and accessible, thanks to the gradual enhancement of computational capabilities, development of new force fields, improved algorithms for polymer structure generation, refined simulation protocols, and advanced analysis tools. Unfortunately, molecular simulations remain resource-intensive and are far from being a black-box approach suitable for routine screening.

Machine Learning (ML) approaches, already widely used in the quantitative structure–activity relationships (QSAR) modeling,[28] prove to be highly effective in predicting polymer properties as well.[29] Broadly, a supervised ML model functions similarly to fitting functions, akin to parametric equations of state. However, ML models trade a solid physical foundation for flexibility and the ability to uncover and utilize previously unknown correlations, if any, between the property to be predicted and any predictor (feature) or set of predictors. Significant research has been conducted using Machine Learning, particularly in its Deep Learning (DL) subfield, to predict sorption in polymers, vapor–liquid equilibriums, *etc*.[30–36] Most notably, DL models can potentially bridge the gap between detailed molecular structure and macroscopic properties, a gap that large-scale MD simulations are partly designed to address.‡ This capability arises from DL models' proficiency in processing diverse data types and identifying complex correlations. Ideally, these models would learn to use molecular structures directly as input, thereby eliminating the need for feature engineering by the researcher. This leads to the critical question of how to represent molecular structures, with existing methods including sequences, graphs, grids, and more.[37–40] In our current research, we adopt the approach described by Casey *et al.*[41] which utilizes electron density and electrostatic potential distributions from quantum-chemical calculations in the form of 3D images as input for a convolutional neural network (CNN). The CNN acts as a feature extractor, learning an optimal representation of molecular structures as feature vectors. These learned features are then fed into a dense neural network (DNN) for making predictions.

Advantages of the electron density distribution as a feature:

(1) **Generalizability**: At a sufficient level of generalization, electron density should enable the use of the same network weights for chemical groups or bond types previously unlearned by the model. In contrast, neural networks that accept more compact molecular representations and generate, for example, electron density, typically need to learn new elements, bonds, or groups beforehand.[42–45]

(2) **Versatility**: As per the seminal Hohenberg-Kohn theorems,[46] electron density alone is sufficient to compute most molecular properties.

(3) **Shapes and volumes**: Real-space density images simplify accessing the volume and shape of chemical entities. Volumetric properties are crucial in various phenomena, including sorption in glassy polymers, diffusion in dense media, and chemical reaction rates.

(4) **Implementation**: There are well-established and widely implemented techniques for handling images, such as the Convolutional Neural Networks (CNNs).

Drawbacks of the electron density distribution as a feature:

(1) **High fitting cost**: CNNs are computationally demanding, leading to time-consuming model fitting. This issue can be partly mitigated by exploiting the sparsity of electron density distribution images and applying existing techniques.[47–51]

(2) **High data computational cost**: The input data, derived from time-consuming quantum-chemical calculations, poses a significant computational challenge. However, advances in hardware and computational techniques are reducing this issue. For example, generating a single electron density distribution image for many polymers is now feasible within minutes using consumer-grade hardware and approximations like GFN2-xTB[52] and RI.[53]

(3) **High memory demands**: The input data is memory-intensive, requiring substantial storage capacity.

(4) **Limited extendability**: The fixed spatial frame and resolution of electron density images may not accommodate larger molecules, such as certain polymer repeating units. Reducing image resolution, as typically done in the standard CNN applications, is not a suitable solution. Consequently, larger frames needed to include all relevant chemical species lead to significantly increased computational demands.

(5) **Normalization challenges**: The values of electron density distribution vary widely, with densities near heavy nuclei being much higher than those near lighter nuclei. This variation poses challenges for NNs. Typical solutions, such as clipping extreme values and scaling the rest,[41] are used, but the effects of such scaling are not well-researched.

In this pilot study, we aim to exploit the aforementioned advantages and establish a connection between routinely computable molecular structures and macroscopic solubility. This is achieved by representing polymers and solvents through the electron densities of their elementary units.

‡Neither Deep Learning (DL) nor Molecular Dynamics (MD)/Monte Carlo (MC) methods are universally superior; each has its niche. DL excels in screening due to its speed but struggles with unlearned patterns, while MD/MC offers detailed insights but is more labor-intensive. The fields are merging, with machine-learned force fields being increasingly used in molecular simulations.

2. Methods

2.1 Model input overview and rationale

Each experimental data point describes the weight fraction (ω_s) of a solvent/gas/vapor in a polymer, characterized by its

molecular weight (M_w), at a specific temperature (T) and gas pressure (P). A polymer and a solvent are represented by the 3D images of their electron density distributions. Electron density images of a single molecule of a solvent, and a single repeating unit of a polymer, are used in each input of the model. The construction of training examples is shown in Fig. 1.

M_w, P, T, and ω_s are the table data collected from the literature as described in section 2.2. The electron density 3D image generation workflow is described in detail in section 2.6. Here we will only outline the main features.

We represent a polymer with a single repeating unit. An appropriate repeating unit is any chemical structure that, when repeated, produces the complete polymer chain and fits within the input image frame. Suppose [–R–] is a minimal repeating unit. Then, two consecutive minimal units [–R–R–] also constitute a valid repeating unit. Regardless of the repeating unit chosen – whether [–R–], [–R–R–], [–R–R–R–], or any other valid unit – we should get the same predictions from the model. For better generalization, we perform data augmentation by using several repeating units to represent a polymer, thereby increasing the training examples for each experimental data point. For instance, representing polypropylene with both [–$CH_2CH(CH_3)$–] and [–$CH_2CH(CH_3)$–$CH_2CH(CH_3)$–] structures means each polypropylene experimental data point will give rise to at least two corresponding training examples.

Next, we should ensure conformational invariance. In 3D space, a molecule is specified by a set of relative positions of atoms. A conformer is a distinct set of atom positions corresponding to a local minimum of the molecule's potential energy. Most of the chosen polymer repeating units have several energetically accessible conformers. These conformers should lead to similar outputs from the model. Hence, for better generalization, we perform data augmentation using as many of these conformers as possible. For instance, suppose we've found 3 conformers of the [–$CH_2CH(CH_3)$–] polypropylene repeating unit and 10 conformers of the [–$CH_2CH(CH_3)$–$CH_2CH(CH_3)$–] repeating unit. Consequently, each polypropylene experimental data point will have at least 3 + 10 corresponding training examples.

The same logic applies to solvent molecules, except there is no need to define different repeating units, only conformers. Consider an experimental data point describing the weight fraction of *n*-butane in polypropylene. Suppose we have identified two conformers of *n*-butane. Now, for each of the different conformers of two polypropylene repeating units, we can use either conformer of *n*-butane. Consequently, each polypropylene + *n*-butane experimental data point will have $3 \times 2 + 10 \times 2$ corresponding training examples.

Finally, to ensure translational and rotational invariance, a standard augmentation technique involves performing various rotations and displacements on the input images. We compared two augmentation methods: one with random on-the-fly shifts and rotations, and another using a series of pre-rotated and pre-shifted images (24 rotated-displaced images plus the original image, achieved through 90-degree rotations about different axes, see Casey *et al.*[41]). In both cases, we ensured that random displacements kept the densities within the image frame.

For the inference, to make a prediction, the model requires electron density distributions (calculated using DFT) for a single arbitrary conformer of a gas molecule and a single arbi-

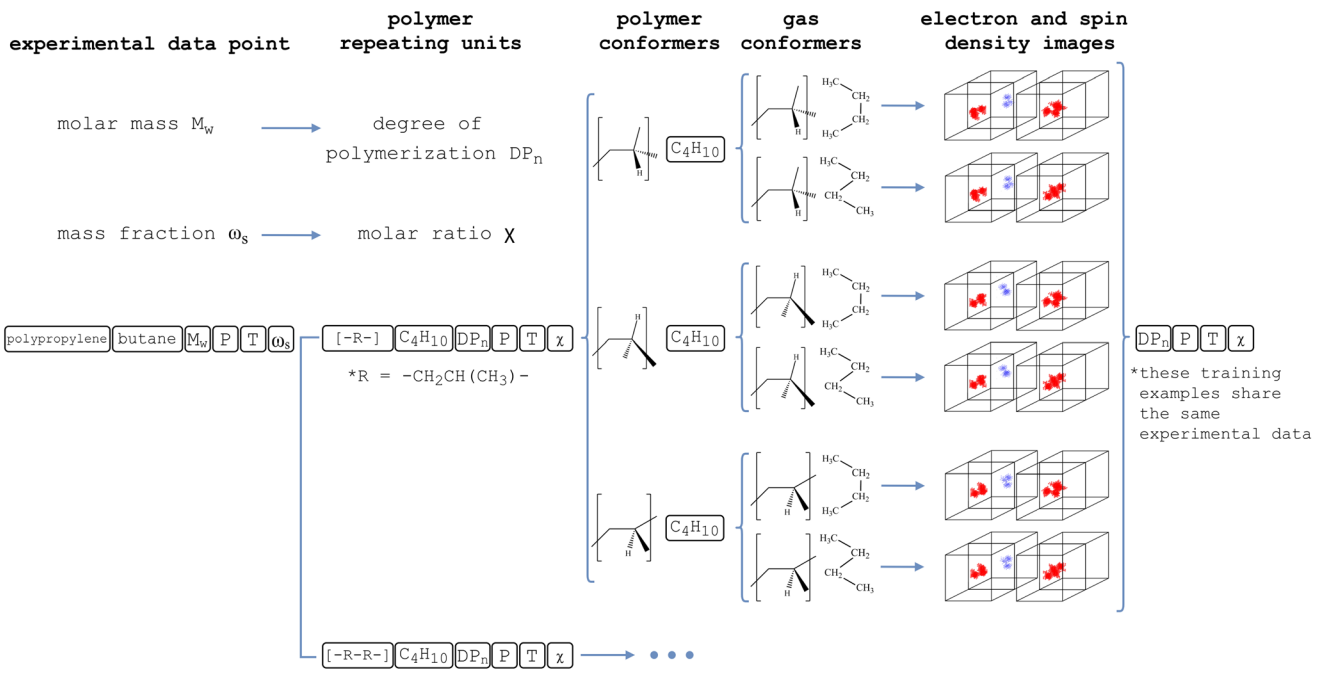

Fig. 1 Training examples assembly scheme. Experimental parameters M_w, P, and T constitute the tabular data. The corresponding electron densities of chemical entities are represented as 3D images. For a detailed explanation, refer to section 2.1.

trary conformer of an arbitrary polymer's repeating unit. The data augmentations previously mentioned are designed to mitigate concerns about selecting specific polymer repeating units, chemical entity conformers, and proper image orientation. However, from a practical standpoint, it is preferable to select the minimal polymer repeating unit and centralize the density in the feed image.

2.2 Experimental macroscopic features

ω_s (unitless): equilibrium (in a kinetic sense) mass fraction of a solvent (gas, vapor, SCF) in the polymer (solid/liquid) phase, the target value;

P (MPa): solvent (gas, vapor, SCF) pressure within a sorption experiment, a macroscopic feature;

T (K): system temperature within a sorption experiment, a macroscopic feature;

M_n and M_w (g mol^{-1}): the number and weight average molar masses of the polymer, respectively, macroscopic features. Only M_w is used in the present study.

Macroscopic features such as the polymer glass transition temperature (T_g, K), density (ρ, g cm^{-3}), crystallinity (X, dimensionless fraction), solvent boiling temperature (T_b, K), critical temperature (T_c, K), and critical pressure (P_c, MPa) are included in the dataset available in the ESI,† but are not used in this study. For further discussion, refer to section 4.1.

2.3 Experimental data limitations

Data range: $0.0001 < \omega_s < 0.5$; $P > 0.0001$ MPa.

State of matter: the solvent is in the gas or supercritical state; the polymer is in the glassy amorphous, rubbery, or molten state under experimental conditions. Although the dataset includes some semi-crystalline polymers, they are not currently utilized.

Polymer structure: only linear polymers with repeating units connected with single bonds are used. Branching is avoided as much as possible. Cross-linked polymers are not considered. Where applicable, polymers are implicitly assumed to be atactic, with atactic variants preferred over stereoregular ones. However, the polymers' tacticity is not specifically addressed unless they exhibit some degree of crystallinity.

Polymer characterization: a polymer should be characterized by its M_n and M_w, or it should be possible to estimate M_w through manufacturer data, typical synthesis conditions, *etc*. For crystallizable polymers, an explicit crystallinity estimation should be available.

Data selection: only the values with a relative experimental uncertainty of less than 10% are collected. In the absence of explicit uncertainties, estimates are roughly made using equipment specifications. Sorption data is prefered over desorption data when both are available. Virgin polymer data are generally preferred, except when treatment is necessary for purposes such as purification or crystallinity elimination. Data presented solely in volume fractions were not collected due to ambiguity.

2.4 Experimental data conversion

The data from the literature were transformed into weight fractions (ω_s) of the solvent in the polymer phase and solvent pressure (P, MPa) as follows:

Solvent amount format: sorption data in cm^3 (gas, STP)/cm^3 (polymer) format were converted using gas densities at STP from NIST[54] and polymer densities provided by the authors. For solvents in the liquid/solid state at STP, a molar volume of 22 414 cm^3 mol^{-1} was employed to calculate their gaseous density at STP. If the polymer density (ρ) was not available within the paper, a tabulated value was used. Polymer swelling was neglected unless explicitly mentioned by the authors. We carefully distinguished between weight/mass fractions defined as m_{gas}/m_{pol} and $m_{gas}/(m_{gas} + m_{polymer})$, with the former being converted to the latter.

Pressure format: pressure (P) data in the form of activities or fugacities were converted to MPa using the models and parameterizations mentioned by the authors. If such details were not specified, activities were assumed to be equal to $P/P_{saturated}$, with P being the actual vapor pressure and $P_{saturated}$ being the saturated vapor pressure. Unless provided by the authors, saturated vapor pressures were taken from NIST or calculated using coefficients from KDB.[55]

Graphical data: data available exclusively in graphic form was digitized using WebPlotDigitizer.[56] Rotation correction was consistently applied during axis alignment. All points were manually placed on the highest quality images available.

Synthetic data: the dual-mode sorption model parameters, when available and reasonable, were used to generate low-pressure sorption isotherms. Occasionally, dual-mode sorption parameters were obtained by fitting data from multiple sources.

2.5 Experimental data correctness check

The following strategy was adopted to check the dataset for outliers and errors that may have occurred during data collection: initially, the entire dataset, including validation and test subsets, was used to fit several simplified and overly regularized models. Subsequently, the examples with the largest relative errors were examined for typos and experimental methodology flaws and compared with similar data from other sources. Typos were corrected, and dubious sources were excluded. If no apparent problems were identified, the data was deemed correct yet challenging, and no further action was taken.

2.6 Electron density distribution images

Polymers and solvents were represented by their 3D electron density images (80 × 80 × 80 voxels, with each voxel measuring 0.4 × 0.4 × 0.4 Å). A single molecule's electron density was used to describe a solvent, while a repeating unit's electron density was used for a polymer. The dangling bonds of the repeating units were not capped. Instead, a triplet state was assigned during the DFT calculation. Consequently, polymer images have two channels: the first containing electron

density and the second indicating the junction points of the repeating units with spin density. The general computational workflow is as follows:

1. For each polymer, several structures of repeating units were chosen as discussed in section 2.1.

2. The dangling bonds of the polymer repeating units are caped with methyl groups (to be removed at the final stages)

3. Resulting molecules are processed using RDKit[57] to generate up to 500 conformers per molecule. These conformers are generated using the ETKDGv2 method[58] and optimized with the MMFF94 force field.[59]

4. The geometries of conformers are optimized using the GFN2-xTB[52] method. The ORCA[60] program package is used for all quantum-chemical in this research.

5. Geometry optimization often leads to duplicate structures, which are removed using the conformers script.[61] In this script, duplicate elimination involves aligning two molecules along the axes of their inertia tensors, calculating their RMSD, with atom correspondence determined by spatial proximity, and renumbering the second molecule if it closely resembles the first. The threshold RMSD is set at 0.2.

6. The conformers' geometries are optimized at the r^{2}SCAN-3c[62] level. A structure with minimal energy is selected within each group of conformers, and those above the 5 kJ mol^{-1} energy window are filtered out. Manual adjustments were occasionally necessary for molecules with several bulky side groups, as non-covalent interactions favor stacked side groups, reducing the number of possible conformers.

7. The caping CH_3-groups in polymer repeating units are removed, and duplicate structures are again eliminated.

8. Single-point calculations at the B3LYP/def2-TZVP[63] level are performed for all conformer structures.

9. High-resolution spin and electron density images in the Gaussian cube-format are generated after single-point DFT calculations. These images are then interpolated on a grid with the same physical frame size as the target but with a higher (typically ×6) resolution. The resolution is subsequently decreased to the target one (80 × 80 × 80 voxels) using summation instead of averaging/interpolation so that the resulting values represent a (scaled) total amount of the electron charge within a voxel's boundaries rather than a value of the interpolated density function at some point. Hence, the sum over voxels is the number of electrons in a molecule, scaled by a common constant for the entire dataset.

2.7 Scaling

The mass fraction of a solvent (gas, vapor, SCF) in a polymer (ω_s) is converted to the molar sorption ratio (χ, $\mathrm{mol_{gas}/mol_{polymer}}$). Unlike ω_s, χ depends on the choice of the polymer's repeating unit and typically ranges from 3×10^{-5} to 1×10^{1}. To reduce the asymmetry of the distribution (Fig. 2) and to avoid overemphasizing strongly sorbing systems, the decimal logarithm $\log_{10}(\chi)$ was used, falling within the range of (−4,1). The weight average molar mass (M_w) of the polymer is converted to the degree of polymerization (DP_n) and used in the logarithmic form. DP_n depends on the choice of the polymer's repeating unit. Pressure (P) is also used in the logarithmic form.

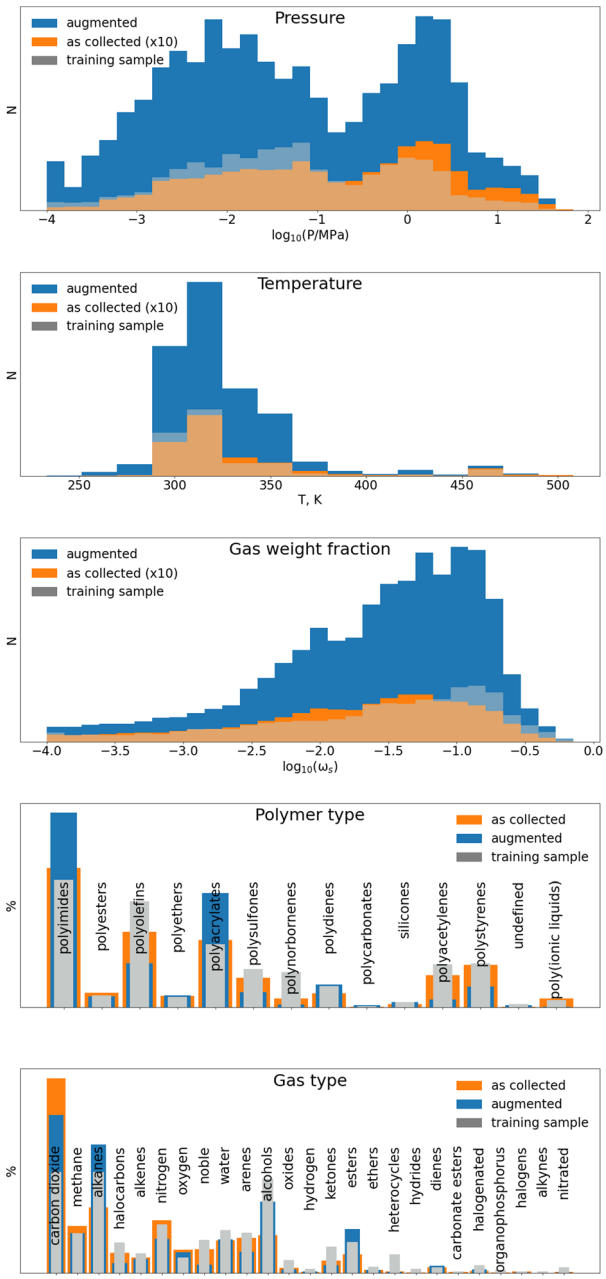

Fig. 2 Distributions of pressure (logarithmic scale), temperature, gas weight fraction (logarithmic scale), polymer types, and gas types for the as-collected experimental dataset (orange), augmented with conformers dataset (blue), and training sample (gray).

Temperature (T), $\log_{10}(M_w)$, and $\log_{10}(P)$ were shifted and scaled to fit within the range (0,1). Standardization was not applied, as the data distributions were not strictly normal (Fig. 2). Electron density images were divided by an empirically chosen coefficient, set at 20, without subsequent clipping.

2.8 Dataset overview

The dataset comprises 14 808 experimental or derived data points from 257 sources,[17,64–318] featuring 102 polymers and 79 solvents/gases, which together form 497 polymer–gas pairs.

A total of 256 unique chemical compounds were considered, accounting for different repeating units of the same polymer, with 2462 unique conformers. The combination of all available augmentations results in approximately 500 000 unique training examples. The version of the dataset used in this research is available in the ESI.† The most recent version of the dataset can be found on GitHub.[319]

2.9 Dataset balancing

Before each epoch, a training subset was randomly selected from the entire dataset to ensure the following conditions were met:

- each polymer/solvent pair had an equal share within the sampled subset;

- experimental data points were equally represented within their respective polymer/solvent pair shares.

Furthermore, the sampling weights of the experimental data points indicating high sorption rates were slightly increased (Fig. 2). The rationale behind this approach is detailed in section 4.2.

2.10 Model

Similarly to the model by Casey *et al.*,[41] our network performs two steps to process spatial electron density distribution (Fig. 3): feature extraction using a convolutional neural network (CNN) and target value prediction *via* a dense/fully connected network (DNN). The feature extraction module of our model employs a network architecture (Fig. 3, left) that is formally akin to certain face recognition networks.[320] The same CNN is utilized to extract both polymer and gas features. Sharing the CNN weights aims to significantly reduce the number of model parameters, mitigate overfitting, and enhance generalization. Ultimately, the feature vectors of the polymer and gas are concatenated with a vector of experimental conditions and then inputted into the DNN.

The CNN block (Fig. 3, right) draws inspiration from the Inception architecture.[321] In each of its four basic blocks, $3 \times 3 \times 3$, $5 \times 5 \times 5$, and $7 \times 7 \times 7$ filters are applied in parallel to the input. Then, computation results are concatenated with the input (forming a residual connection) and fed into a $1 \times 1 \times 1$ layer for dimension reduction. Additionally, we experimented with basic CNN blocks similar to those used by Casey *et al.*,[41] as well as convolutional networks akin to the DenseNet model,[322] employing only $3 \times 3 \times 3$ kernels in both instances. All types of neural networks successfully minimized the loss, but the model with multiple kernels required significantly fewer parameters to achieve this. Generally, a more targeted selection of model details is beyond the scope of this work.

After each basic block, the output is downsampled using average pooling with a $2 \times 2 \times 2$ window and a shift of two voxels, followed by element-wise multiplication by 8. Effectively, sum pooling is performed. Thus, the vanishing of the sparse input channels through repetitive average pooling is mitigated. The original meaning of the input voxels (total elec-

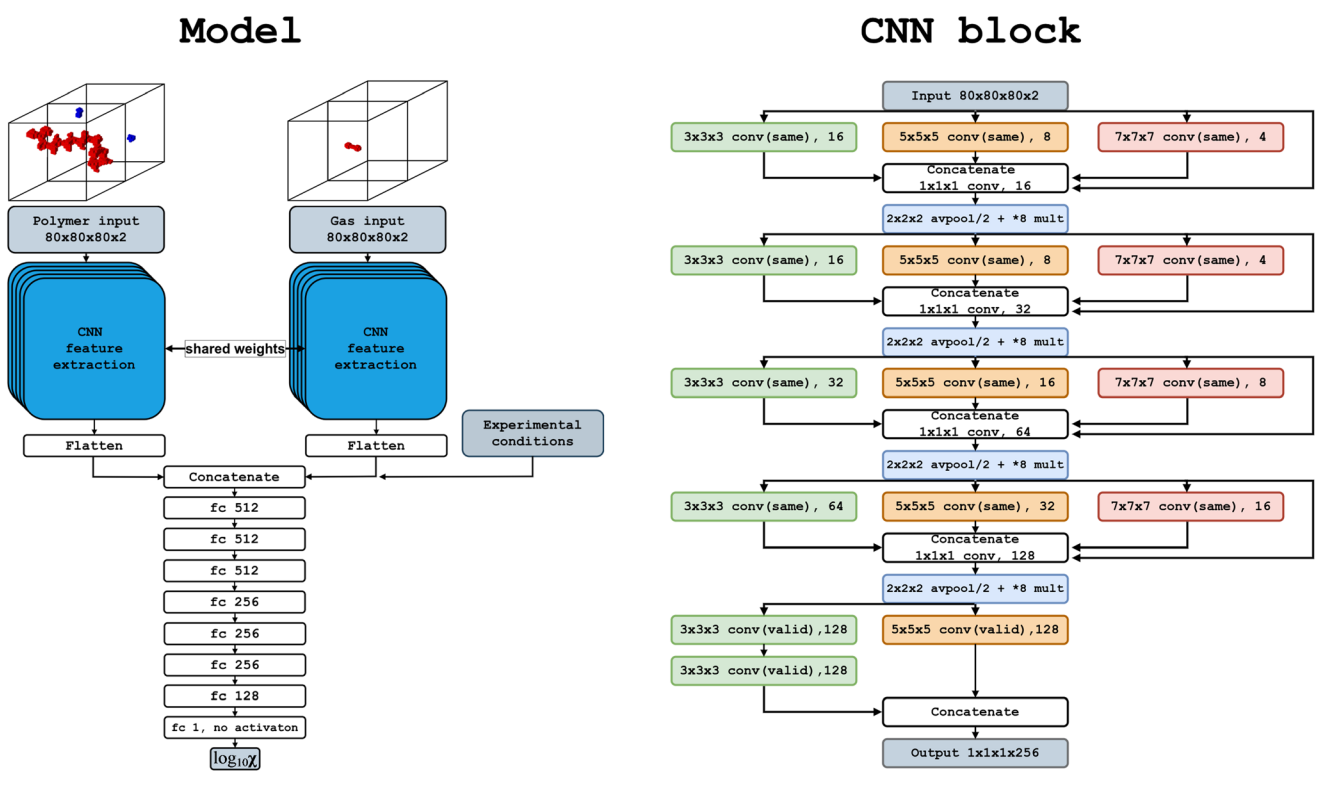

Fig. 3 The model architecture (left) and the convolutional network used for feature extraction (right). For a detailed explanation, refer to section 2.10.

tron charge within a voxel) is preserved at the expense of the additional non-linearity provided by max pooling. No systematic comparison with max pooling was conducted. Both DNN and CNN blocks employ ELU activation. The model comprises 5 million parameters, with the CNN block containing 3.6 million parameters and a FLOP count of 33.31 billion. The model is implemented using TensorFlow 2[323] and Keras[324] and is available on GitHub.[319]

2.11 Fit and test

The mean squared error (MSE) between $\log_{10}(\chi)$ (experimental) and $\log_{10}(\hat{\chi})$ (predicted), where χ is the molar sorption ratio ($mol_{gas}/mol_{polymer}$), was employed as the loss function. The Adam optimizer was utilized for optimization with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and AMSGrad set to False. The learning rate was set to 5×10^{-6} as higher rates tended to cause optimization instability. The minibatch size was fixed at eight examples. L2 regularization was applied, utilizing a weight decay parameter λ of 5×10^{-5}. The training procedure was executed using mixed precision (automatic mixed precision, AMP), employing both 16-bit and 32-bit floating-point types for acceleration. The target irreducible mean absolute percentage error (MAPE) for ω_s estimation was set at 20%, corresponding to a mean absolute error (MAE) of approximately 0.085 and a mean squared error (MSE) of 0.013 for $\log_{10}(\chi)$ estimation. For justification of the target error rate, refer to section 4.3. If the electron density images were augmented on the fly, the shift and rotate functions from the ndimage submodule of the SciPy module[325] were applied to generate randomly shifted and rotated images, setting the spline interpolation order to 3.

For training, the dataset was randomly divided into training, validation, and test sets based on polymer–solvent pairs, ensuring that no polymer–solvent pair appeared in more than one set simultaneously. The division was approximately in a 70/15/15 ratio. Additionally, the splitting process ensured that no polymer from the test set appeared in the training or validation sets. While the training and validation sets partially shared the same polymers, they contained different conformers of the shared polymer repeating units. The procedure aimed to maintain a proportional representation of different chemical types of polymers (*e.g.*, polyolefins, polyimides) and solvents (*e.g.*, simple compounds, olefins). A subset of 176 192 examples was randomly sampled from the training set before each epoch, as discussed in section 2.9. Conversely, a fixed subset of approximately 22 976 examples for validation was sampled from the validation set only once at the beginning of the fitting procedure.

The test set included sorption data for 22 polymers (Fig. 4), all previously unseen by the model. No special attention was given to ensuring that the test set contained chemical groups or elements not encountered by the model during training. However, the test set does occasionally include unique UNIFAC groups[326,327] absent from the training and validation sets. For instance, the aromatic fluoro group ([38] ACF) in 6FDA-OFB (id 76) and the aromatic chloro group ([25] ACCL) in 6FDA-5CMPD (id 77) are examples. The experimental parameters within the test set span a wide range, covering cases from the sorption of supercritical fluids (SCFs), such as supercritical CO_2 under typical SCF application conditions, to the sorption of vapors, including volatile organic compounds (VOCs), and the low-pressure sorption of gases like CH_4, N_2, and O_2, common in membrane separation conditions.

3. Results

3.1 Translational invariance and model performance

The dataset was augmented with shifted and rotated electron density images using two methods: on-the-fly random image shifts and rotations before every training batch was fed to the model; and a pre-augmented dataset containing pre-rotated and pre-shifted images, which comprised 24 rotated and displaced images in addition to the original image, achieved through 90-degree rotations about different axes. Over 150 epochs, the model, employing either dataset augmentation method, seemingly converged with a mean absolute error (MAE) for $\log_{10}(\chi)$ of approximately 0.065 (Fig. 5), close to our estimated irreducible MAE of 0.085. The model underwent further training for an additional 150 epochs (not shown) with the pre-augmented dataset, resulting in only marginal accuracy improvements and no signs of overfitting. The validation MAE for $\log_{10}(\chi)$ was 0.22 ± 0.02 for the pre-augmented dataset and 0.17 ± 0.02 for the on-the-fly augmented dataset, with the ±0.02 interval indicating the oscillation range of about 0.04 in the validation MAE.

First, we need to evaluate the effectiveness of two data augmentation options in approximating rotational and translational invariance. To measure how closely the model achieves invariance, we will use a modified mean absolute deviation (MAD) around the mean. During both training and testing, the model processes multiple examples from the same experimental data point and the same polymer repeating unit and gas molecule conformers. These examples differ only in the position and orientation of the electron density within the image frame. For each combination of an experimental data point and conformers' choice, we will calculate the individual $\log_{10}(\chi)$ MAD_i around its individual mean: $MAD_i = \frac{\sum_{displacements} \left| \hat{y} - \bar{\hat{y}} \right|}{N_{displaced}}$, where $y = \log_{10}(\chi)$, and $N_{displaced}$ represents the number of examples related to the same experimental data point, involving identical polymer repeating unit's and gas molecule's conformers, but differing solely in image orientation. We will then compute the average of these individual MADs over the entire test set: $MAD(displacements) = \overline{MAD_i}$. This approach is valid as $\log_{10}(\chi)$ normalizes sorption rates of varying magnitudes to a comparable scale.

Using this definition, we obtained the following MAD(displacements) estimations: 0.02 for on-the-fly image augmentation and 0.06 for the pre-augmented dataset. On-the-fly image

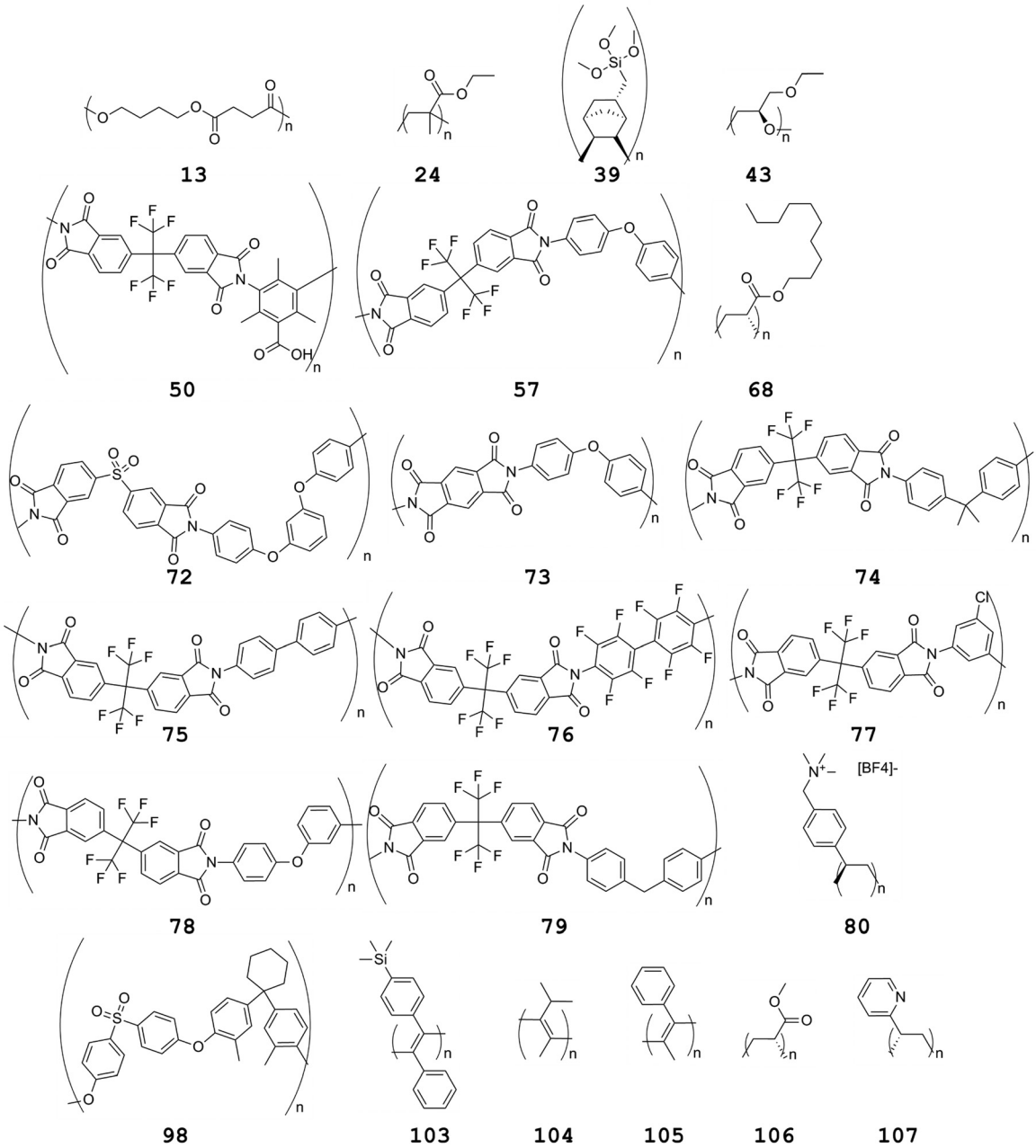

Fig. 4 Polymers included in the test set (the numbers correspond to IDs from the dataset provided in the ESI†). **13**: poly(butylene succinate), **24**: poly(ethyl methacrylate), **39**: poly(5-methyltrimethoxysilyl-2-norbornene), **43**: poly(ethyl glycidyl ether), **50**: 6FDA-TrMCA, **57**: 6FDA-4,4′-DPE, **68**: poly(decyl acrylate), **72**: DSDA-TPER, **73**: PMDA-4,4′-DPE, **74**: 6FDA-6HPDA, **75**: 6FDA-OHB, **76**: 6FDA-OFB, 77: 6FDA-5CMPD, **78**: 6FDA-3,4′-DPE, **79**: 6FDA-4,4′-DDM, **80**: poly(*p*-vinylbenzyltrimethyl ammonium tetrafluoroborate), **98**: 3,3′-dimethylbisphenol-Z polysulfone, **103**: poly(1-phenyl-2-[*p*-(trimethylsilyl)phenyl]acetylene), **104**: poly(4-methyl-2-pentyne), **105**: poly(1-phenyl-1-propyne), **106**: poly(methyl acrylate), **107**: poly(2-vinylpyridine).

augmentation allows for a satisfactory approximation of rotational and translational invariance, as the variations of the predictions introduced by spatial transformations of electron density images are minor and comparable to the oscillations of validation MAE. On the other hand, the pre-augmented dataset was not sufficient to achieve invariance. The model trained on the pre-augmented dataset will not be further considered. All subsequent results refer to the model that was trained using on-the-fly augmentation.

Test MAE($\log_{10}(\chi)$) was determined to be 0.15, corresponding to a MAPE(ω_s) of 32% (Table 1). The test MAE falls within the range of validation MAE variations observed during fitting (0.17 ± 0.02). Fig. 6 presents illustrative examples of the model's performance, and an overview of the entire test set is provided in Table 1.

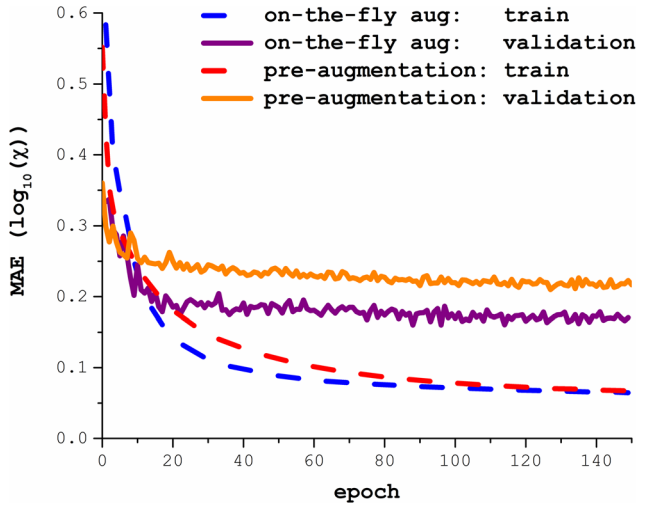

Fig. 5 The history of the mean absolute error (MAE) between the experimental $\log_{10}(\chi)$ and predicted $\log_{10}(\hat{\chi})$ throughout the training of the model. The model employed two augmentation methods: on-the-fly random image shifts and rotations, and pre-augmented images, which underwent 24 different 90-degree rotations and shifts about various axes. Here, χ denotes the molar sorption ratio ($mol_{gas}/mol_{polymer}$). Approximately 176 000 examples (equivalent to about 22 000 minibatches) were utilized for fitting in each epoch, and roughly 23 000 examples were used for validation.

3.2 Error analysis

To achieve conformational invariance, the dataset was extensively augmented with different conformers of polymer repeating units and gases. To assess the model's invariance, similar to the approach used above for each experimental data point, we calculated the individual $\log_{10}(\chi)$ MAD_i around its individual mean: $MAD_i = \frac{\sum_{conformers} |\hat{y} - \bar{\hat{y}}|}{N_{conformers}}$, where y is $\log_{10}(\chi)$, averaged in advance over all feed images' displacements and rotations. Subsequently, we computed the mean of these individual MADs across the entire test set: $MAD(conformers) = \overline{MAD_i}$. Using this definition, we estimated MAD(conformers) to be approximately 0.02, which, again, is comparable to the oscillations of the validation MAE and indicates the satisfactory (on average) approximation of conformational invariance.

There are two specific situations where conformational invariance is not achieved. The first involves polymers with long and flexible substituents, such as poly(ethyl glycidyl ether) (PEGE, id 43) and poly(decyl acrylate) (PDA, id 68). Their MAD(conformers) values are 0.1 and 0.16, respectively, with the largest variance corresponding to the polymer with the longest (decyl) substituent (Fig. 7). In the latter case, the failure to achieve invariance significantly impacts the model's accuracy. This indicates a deficiency rooted in the complex phase behavior of polymers with lengthy side chains. While long substituents, such as pendant alkyl chains, can form something like a distinct 'phase' with its own crystallization behavior and melting point, the properties of such a 'phase' are still markedly different from those of the corresponding idealized phase, *e.g.*, polyethylene.[194] Moreover, this behavior significantly diverges from that of most single-phase polymers the model has encountered during training.

The second scenario of variation associated with conformer choice involves the possibility of specific interactions between a sorbed gas and a polymer. For example, the test set includes data on water sorption in poly(ethyl methacrylate) (PEMA, id 24), 6FDA-4,4′-DPE (id 57), DSDA-TPER (id 72), PMDA-4,4′-DPE (id 73), and poly(2-vinyl pyridine) (P2VP, id 107). The model's predictions for polymers containing polar or hydrophilic substituents, namely PEMA and P2VP, depend on the conformer choice, with MAD(conformers) values of 0.14 and 0.37, respectively, for PEMA/H_2O and P2VP/H_2O pairs (Fig. 7). Generally, for the rest of the solvents, the model's predictions for PEMA and P2VP appear to be conformationally invariant, with MAD (conformers) values of 0.03 in both cases. Similarly, the model's predictions for water sorption in polymers with no polar substituents (6FDA-4,4′-DPE, DSDA-TPER, PMDA-4,4′-DPE) do not seem to depend on the choice of a conformer. However, the lack of invariance is associated with significant errors. For example, in the case of water sorption in PEMA, the model's prediction MAE($\log_{10}(\chi)$) was determined to be 0.26, corresponding to a MAPE(ω_s) of 91%, while for the rest of the gases sorbed in PEMA, these values were 0.15 and 35%. It is worth noting that the model has thus acquired some chemical intuition. Nevertheless, this remains a complex issue that likely requires resolution through the collection of a larger training set and careful attention to the quality of water sorption data.

Another significant source of deviation between the model's predictions and experimental results arises from polymer–gas pairs involving polar solvents, including butyl acetate (MAE($\log_{10}(\chi)$) 0.36, MAPE(ω_s) 131%), ethyl acetate (MAE($\log_{10}(\chi)$) 0.34, MAPE(ω_s) 121%), propyl acetate (MAE ($\log_{10}(\chi)$) 0.25, MAPE(ω_s) 77%), and methanol (MAE($\log_{10}(\chi)$) 0.43, MAPE(ω_s) 59%). Notably, excluding their sorption in poly (methyl acrylate) (PMA, id 106) results in a significantly reduced MAE($\log_{10}(\chi)$) of 0.13, accompanied by a corresponding MAPE(ω_s) of 33%. However, the model demonstrates good performance in predicting the sorption of other polar solvents, such as methyl acetate and methyl ethyl ketone. Consequently, it remains challenging to definitively ascertain whether such solvents inherently pose a challenge to the model, potentially involving specific interactions between the polymer and the solvent, or if the observed discrepancies are attributed to experimental data peculiarities, including characteristic measurement methods such as saturated vapor pressure and potential impurities like water.

Finally, as will be discussed in section 4.3, sorption in certain polymers is highly dependent on the sample preparation protocol and sample history. This is particularly true for polyacetylene-type polymers like poly(4-methyl-2-pentyne) (PMP, id 104) and poly(1-phenyl-1-propyne) (PPP, id 105). In these polymers, the variability in measurement results due to different sample histories can reach up to 50%. Therefore, dis-

Table 1 Polymers included in the test set and corresponding model performance

Polymer	Polymer abbr., dataset ID	Solvents	Exp. points (no. of sources)[a]	ω_s MAPE, % ($\log_{10}(\chi)$ MAE)[b]
Poly(butylene succinate)	PBS, 13	CO_2	39 (1)	13.2 (0.061)
Poly(ethyl methacrylate)	PEMA, 24	CO_2, CH_4, N_2, CH_2Cl_2, C_2H_4, C_3H_8, Ar, H_2O, $C_6H_5CH_3$, $CH_3C(O)C_2H_5$, $CH_3CO_2C_3H_7$, C_6H_6, $C_6H_5C_2H_5$	340 (8)	40.5 (0.164)
Poly(5-methyltrimethoxysilyl-2-norbornene)	PMeTMeOSN, 39	CO_2, N_2	7 (1)	11.5 (0.056)
Poly(ethyl glycidyl ether)	PEGE, 43	CO_2	20 (1)	13.5 (0.059)
6FDA-TrMCA	50	CO_2, CH_4	12 (1)	16.3 (0.081)
6FDA-4,4′-DPE	57	CO_2, C_3H_6, H_2O, CH_3OH, C_4H_6 (1,3-butadiene)	76 (7)	26.6 (0.116)
Poly(decyl acrylate)	PDA, 68	CO_2, CH_4	8 (1)	120.0 (0.301)
DSDA-TPER	72	H_2O	15 (1)	21.0 (0.113)
PMDA-4,4′-DPE	73	CO_2, H_2O	44 (3)	31.8 (0.154)
6FDA-6HPDA	74	CO_2, CH_4, N_2, O_2	272 (1)	14.0 (0.056)
6FDA-OHB	75	CO_2, CH_4, N_2, O_2	272 (1)	21.6 (0.077)
6FDA-OFB	76	CO_2, CH_4, N_2, O_2	272 (1)	22.3 (0.114)
6FDA-5CMPD	77	C_4H_6 (1,3-butadiene), C_2H_4O (ethylene oxide)	15 (2)	14.3 (0.072)
6FDA-3,4′-DPE	78	CO_2, C_4H_6 (1,3-butadiene)	30 (2)	94.7 (0.241)
6FDA-4,4′-DDM	79	C_4H_6 (1,3-butadiene)	13 (1)	45.3 (0.169)
Poly(*p*-vinylbenzyl trimethyl ammonium tetrafluoroborate)	P[VBTMA][BF_4], 80	CO_2	11 (2)	70.9 (0.557)
3,3′-Dimethyl-bisphenol-Z polysulfone	DMPSF-Z, 98	CO_2, CH_4, N_2	32 (1)	25.8 (0.096)
Poly(1-phenyl-2-[*p*-(trimethylsilyl) phenyl]acetylene)	PTMSDPA, 103	C_3H_8, *n*-C_4H_{10}	28 (1)	82.1 (0.779)
Poly(4-methyl-2-pentyne)	PMP, 104	CO_2, CH_4, N_2, C_3H_8, *n*-C_4H_{10}	53 (4)	56.9 (0.459)
Poly(1-phenyl-1-propyne)	PPP, 105	C_3H_8, *n*-C_4H_{10}	13 (1)	45.5 (0.316)
Poly(methyl acrylate)	PMA, 106	CO_2, $CH_3CO_2CH_3$, $CH_3CO_2C_2H_5$, $CH_3CO_2C_3H_7$, $CH_3CO_2C_4H_9$, C_6H_6	72 (3)	84.1 (0.258)
Poly(2-vinyl pyridine)	P2VP, 107	CH_2Cl_2, H_2O, CH_3OH, C_6H_6	25 (1)	54.2 (0.423)
Average over the test set	—		1669 (36)	32.5 (0.148)

[a] Number of experimentally measured pressure and temperature-dependent gas uptakes, as collected from various literature sources (indicated in brackets). [b] Mean absolute percentage errors (MAPE) between the experimental ω_s (weight fraction) and the predictions made by our model. The corresponding mean absolute errors (MAE) between the experimental values and predicted $\log_{10}(\chi)$ (where χ is molar sorption) are given in brackets.

crepancies between model predictions and experimental data for these polymers are not unexpected. For poly(1-phenyl-2-[*p*-(trimethylsilyl)phenyl]acetylene) (PTMSDPA, id 103), the challenge is compounded as the experimental data comes from a single source, which prevents even partial elimination of the influence of the experimental protocol through averaging. As will be discussed in sections 4.1 and 4.3, more accurate predictions for such polymers could be achieved by considering the macroscopic characteristics of the polymer samples or by incorporating the sample preparation protocol as an additional feature in the analysis.

In general, there appears to be no apparent correlation between the errors in the model's predictions and the experimental conditions (Fig. 8). Regardless of the temperature and actual solvent weight fraction, most of the model's predictions tend to fall within a 50% range. However, there is a noticeable increase in the relative number of large errors as the pressure approaches 0.1 MPa. This trend can be attributed to the fact that data for all the problematic cases mentioned above, such as water sorption, sorption in polyacetylene-type polymers, *etc.*, are predominantly available only in this pressure domain which is specific to membrane science experiments and the studies of VOC (Volatile Organic Compounds) vapor sorption.

As an indication of the model's generalizability, it successfully predicted sorption in polymers with unique UNIFAC groups that were not included during the fitting process. For instance, sorption in 6FDA-OFB (id 76), which contains the aromatic fluoro group, was predicted with MAE($\log_{10}(\chi)$) of 0.11 and MAPE(ω_s) of 22%. Similarly, sorption in 6FDA-5CMPD (id 77), containing the aromatic chloro group, had MAE($\log_{10}(\chi)$) of 0.07 and MAPE(ω_s) of 14%. These results surpass the average test set performance of 32% MAPE. In the subsequent sections, we will provide chemical insights to justify the data handling strategies and feature selections made in this research.

4. Discussion

4.1 Macroscopic feature selection

The dataset, as outlined in section 2.2, excludes most macroscopic features:

(1) Using all macroscopic features often leads the model to persistently default to a local minimum, overlooking electron density data. For such a correlation example see Kanehashi and Nagai.[329]

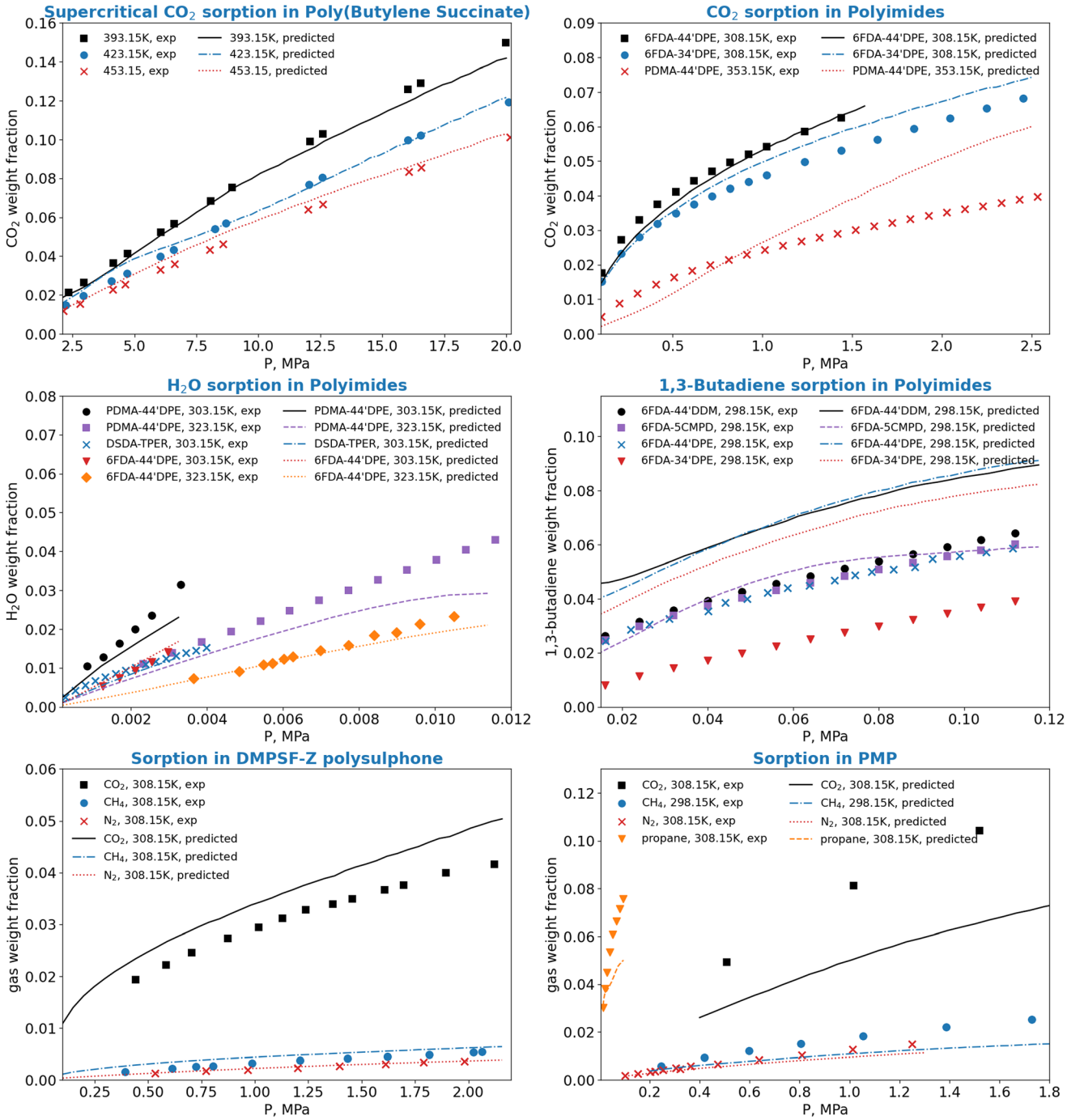

Fig. 6 Predicted (lines) *vs.* exp (dots) ω_s (weight fraction) of gases/vapors sorbed in the polymers from the test set (*i.e.* the polymers unfamiliar to the model). Predictions are made on unshifted electron densities. Supercritical CO_2 sorption in poly(butylenes succinate) (top left),[328] CO_2 sorption in polyimides (top right),[175,180,212] H_2O sorption in polyimides (middle left),[179,243] 1,3-butadiene sorption in polyimides (middle right),[174] sorption of CO_2, CH_4 and N_2 in DMPSF-Z (bottom left),[265] sorption in poly(4-methyl-2-pentyne) (PMP) acetylene-based polymer (bottom right).[277,278,317,318]

(2) Due to the high correlation between M_n and M_w and often only one being accurately known, we've opted to include only M_w in our study.

(3) Polymer density (ρ), though potentially reducing variance related to treatment history (refer to section 4.2), is frequently not measured accurately and substituted by tabulated values in literature, diminishing its significance. We've excluded it from our model but may consider it in future research.

(4) The glass transition temperature (T_g) of polymers, strongly linked with monomers' structure and average molecular weight, is often represented in literature by tabulated

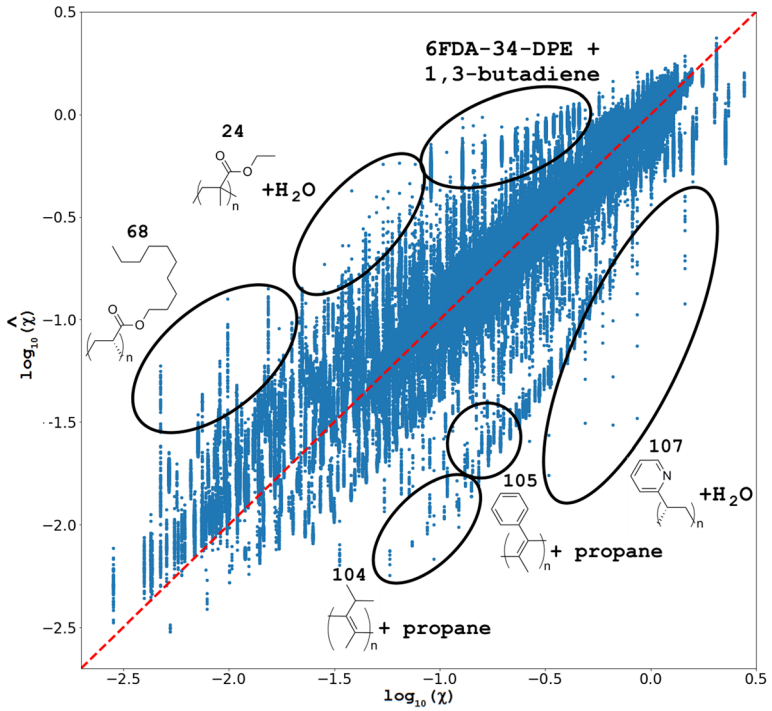

Fig. 7 Predicted $\log_{10}(\hat{\chi})$ *vs.* experimental $\log_{10}(\chi)$ within the test set, with χ being the molar sorption ratio mol_{gas}/mol_{pol}.

values. Due to this, it is omitted from our study, though it could be a useful fine-tuning parameter for understanding polymer treatment history and structural details like tacticity and aggregation.

(5) Solvent critical parameters, primarily determined by the solvent's structure, are also omitted.

(6) Semi-crystalline polymers are excluded due to conceptual issues. In the first approximation, crystallinity eliminates a constant fraction of sorption, even of the lightest gases,[330] and even in many exotic cases.[331,332] This is mostly typical for the sub-T_g domain. The issues are:

(6.1) If the first approximation is valid, the resulting model should only learn to perform a trivial multiplication of the predicted sorption value by a user-defined coefficient (unity minus crystallinity).

(6.2) Crystallinity of the polymer, mostly in the super-T_g domain, may increase (crystallization), decrease, or even vanish (melt). The user should either have an *a priori* knowledge of the polymer crystallization behavior, or the model should be able to predict crystallization behavior. A standalone model for polymer phase behavior prediction could probably be a better solution.

(6.3) In rare cases, a semi-crystalline sample can sorb better than its fully amorphous counterpart.[73] Specific crystallization techniques may produce ordered mesophases exhibiting large free volume.[333]

(6.4) Dividing a semi-crystalline polymer into perfectly impenetrable crystalline and perfectly sorbing amorphous domains is too simplistic.[334] The 'crystallinity degree' is ambiguous,[335] as a real polymer may contain both fully ordered (ideal crystalline) domains, low quality/disordered/small crystalline domains, amorphous domains exhibiting to some extent a short-range order, and finally, fully disordered domains with no clear boundary between the domains. As a result, different methods, such as density measurement, DSC, XRD, SSNMR, *etc.*, will typically provide quite different values of the crystallinity degree. Its estimation will also heavily depend on the approximations used for the raw data interpretation.

Our model employs a minimal set of uncorrelated macroscopic parameters. While this approach reduces the model's accuracy and overlooks certain aspects of polymer morphology and processing history, it lays the foundation for further accounting for the complex nature of polymer behavior. The inherently stochastic nature of polymers may also necessitate a more nuanced application of domain knowledge beyond simply selecting appropriate features, as discussed by Martin *et al.*[336]

4.2 Dataset balancing motivation

The approach implemented in this study to balance the data viewed by the model in each epoch is detailed in section 2.9. Here, we provide a more in-depth explanation and rationale. The dataset is significantly unbalanced due to two main factors:

(1) Polymers and solvents with large, flexible units have more conformers compared to those with small, rigid units. For instance, we included 21 conformers for *n*-nonane, whereas methane has only one. As a result, the dataset contains 21 times more copies of each experimental measurement

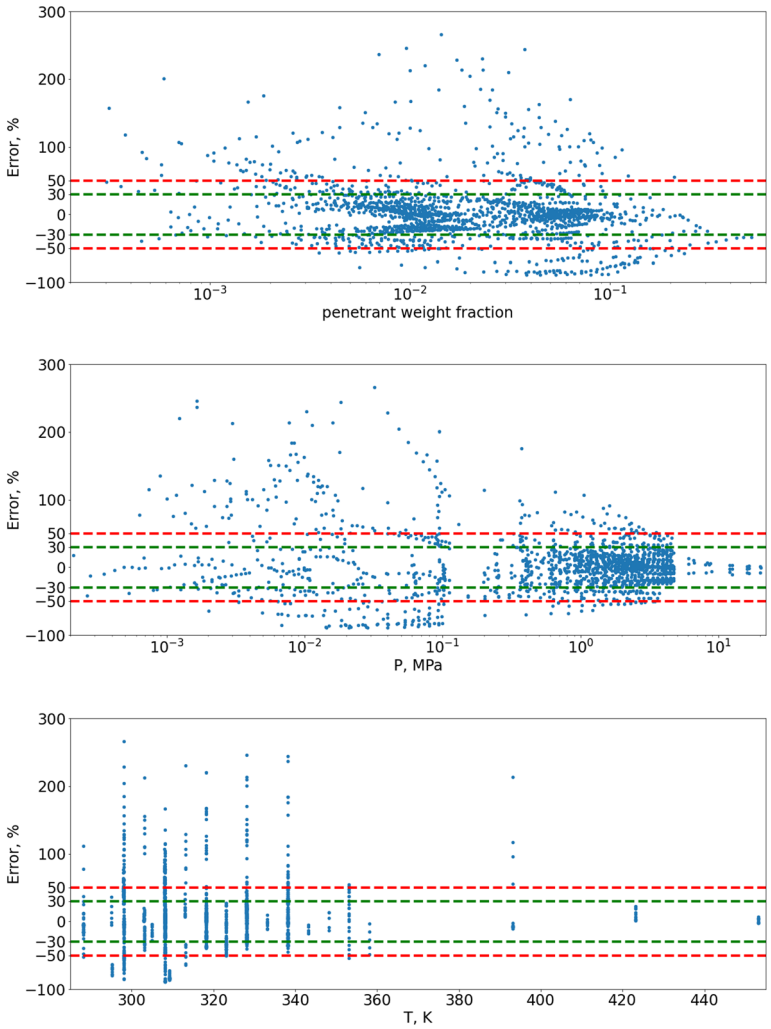

Fig. 8 Signed percentage errors between the experimental ω_s and the predictions made by the model *vs.* ω_s (top), pressure P (middle, logarithmic scale), and temperature T (bottom).

for *n*-nonane than for methane, if the polymer is the same in both measurements.

(2) Experimental data availability is uneven. Certain polymer–solvent systems are extensively studied due to their practical significance (*e.g.*, PEG/CO_2 and 6FDA-DAM/CO_2) or as standard models in developing new techniques and theories (*e.g.*, PIB/benzene, PMMA/CO_2). Data on these systems are abundant, spanning numerous research papers and hundreds of experimental points. However, the dataset includes a wide variety of chemical species for better generalization. Only a few data points are available for less common substances like dimethyl methylphosphonate or SF_6, often sourced from a single study.

To mitigate these issues, we balanced the representation of different polymer–solvent pairs in each training epoch's dataset sample. For instance, the dataset contains approximately 500 experimental data points for CO_2 solubility in PEG. Suppose each electron density image is augmented with 24 of its rotated and shifted copies, totaling 25 images. The augmentation can be performed either in advance or on-the-fly, as discussed in section 2.1. Hence, there can be approximately 9 000 000 training examples concerning CO_2 solubility in PEG (500 measurements × 29 PEG repeating units' conformers × 25 shifted-rotated PEG images × 1 CO_2 conformer × 25 shifted-rotated CO_2 images). On the other hand, there are only five data points for ethylene oxide sorption in 6FDA-TMPPD, leading to approximately 40 000 training examples. Yet, the model processes approximately 500 randomly sampled examples of each polymer–solvent pair per epoch.

However, an inherent imbalance remains due to the uneven availability of data on specific polymers and solvents. For example, data on interactions with 37 different solvents are available for each of PIB and PVAc, while only one polymer–solvent pair exists for the rare poly(trimethylgermyl)propyne. Attempting to balance the training sample by equalizing the representation of various polymer–solvent systems, polymers, solvents, or experimental conditions can inadvertently overemphasize data from single sources. This is problematic, as these

sources might rely on flawed methodologies, potentially leading to inaccurate data and inadvertently increasing the proportion of outliers. We opted to balance polymer–solvent system representations as a compromise between the benefits of dataset balancing and the risks of amplifying outliers. To overcome this issue, a goal to collect as diverse data as possible should be set in future research.

4.3 Irreducible error rate

Estimating the irreducible error rate is crucial in any machine learning research. Predictions from even a perfectly chosen model with ideally fitted parameters will almost always deviate from real-world data, a deviation known as the irreducible error. This error primarily stems from immeasurable variation (mainly noise) and unaccounted variables (*i.e.*, the predicted value might depend on additional parameters not included in the model). Estimating the irreducible error is a complex task. We have estimated the target MAPE to be around 20%, based on a preliminary dataset overview. However, a more comprehensive analysis is planned for the future. Here, we aim to provide insight into the scale and sources of variance. We can identify at least three primary sources of variance: experimental uncertainties, flaws in experimental methodology, and variance related to a polymer's macro- and microstructure. We will briefly discuss these factors, acknowledging that our overview is not exhaustive.

(1) **Experimental uncertainty** is the most controllable and transparent source of variance. Best practice suggests reporting confidence intervals and equipment specifications alongside experimental data. This practice enables straightforward filtering of excessively noisy data once a desired threshold is established. However, confidence intervals are often poorly estimated or omitted, as discussed by Box *et al.*[337] It's also worth noting that temperature or pressure uncertainty, while usually almost negligible, can still be significant. For example, if a sorption isotherm is highly concave towards the sorbed amount axis at high vapor pressures, a minor shift in pressure could result in a substantial change in the predicted value. We have excluded such isotherm tails from our collection.

(2) **Flaws in experimental methodology** can be challenging or impossible to detect. We include here equipment malfunction, unjustified approximations in data processing, misprints in research papers, *etc*. For a comprehensive review of experimental methods and their pitfalls, refer to the literature.[338,339] We will shortly comment on some typical cases:

(2.1) **Equipment malfunction**, loose calibration, failure to correctly collect readings, and poor quality of reagents are some typical issues. These problems are difficult to identify *a priori* in an otherwise well-written research paper. However, data affected by such errors should be excluded. If there is a substantial amount of similar data, a visual inspection of values from multiple sources can help identify and rule out suspicious measurements. It is important, though, to first consider other possible causes of deviation, such as the polymer treatment history, *etc.*

(2.2) **Buoyancy correction** is necessary when using gravimetric methods in dense media like supercritical fluids. If a polymer swells significantly, the buoyancy correction should be adjusted to reflect this change. However, measuring this swelling experimentally is not always feasible. Consequently, various equations of state (EoS) are employed to account for changes in the polymer's volume.[340] Different EoSs can yield significantly different predictions, with both excellent agreement with experimental results and substantial deviations being possible. For instance, a study on CO_2 sorption in polypropylene at high pressures and temperatures[202] demonstrated that incorrect buoyancy correction could lead to an under- or overestimation of sorbed gas by approximately 40–90% at high-pressure limits (~35 MPa). To estimate uncertainty without an in-depth analysis—which would involve considering various EoSs, their parameterizations, and specific experimental conditions—two approaches are viable. One can either use worst-case scenarios to estimate a 50% margin of error or rely on literature estimates, such as the 20% margin suggested by Sato *et al.*[328] These considerations are equally applicable to the pressure decay method.

(2.3) **Spectroscopic methods** (mostly FTIR) are widely used to measure the concentration of a sorbed gas in the polymeric phase.[341] However, the accuracy of quantitative results depends on the knowledge of the gas's molar extinction coefficient. Thus, the approximations regarding molar absorptivity made by authors in research papers should be scrutinized closely. For instance, employing a constant extinction coefficient to the measured band intensity is often incorrect. Conversely, integrating the entire band, dyad, or triad might result in values that are more or less independent of pressure and temperature.[70]

Tracking these issues is a labor-intensive process. It necessitates detailed analysis and evaluation of research papers, rather than mere blind data collection. So far, our approach has primarily followed the method described in section 2.5. A more comprehensive analysis is anticipated in future versions of the dataset.

(3) **Variance related to a polymer's macro- and microstructure** stems from individual features of the polymer samples, such as their treatment history:

(3.1) **Glassy polymers** are non-equilibrium systems. Their sorption characteristics heavily depend on the so-called excess free volume.§ It is a void space conventionally thought to be "frozen" into the polymer matrix and partially accessible to penetrant gases. The sample preparation process (such as solution casting, spin coating, extrusion), specific details (like casting solvent, cooling rate), and pretreatment methods (aging, annealing, quenching, preswelling) significantly influence sorption. Polymer treatment can promote relaxation, thereby reducing free volume. Alternatively, it can create a sparser spatial arrangement of polymer chains and then fix the resulting more loosely packed structure. Given the

§ It will be denoted just as "free volume" for short.

immense chemical diversity, making a general estimation of the spread in sorption values due to free volume variations is challenging, and it is not even clear which state of a particular polymer should be taken as a reference. We will use some representative cases to illustrate the scale of these effects. In a study by Fechter *et al.*,[342] sub-T_g annealing of poly(methyl methacrylate) resulted in a decrease in CO_2 uptake at 700 mm Hg by approximately 13%. Chan and Paul[343] found that sub-T_g annealing of polycarbonate resulted in a 22% reduction in CO_2 uptake at 15 atm and a decrease in the Langmuir capacity constant by 30%. Conversely, preswelling polymer samples increases gas uptake rates. However, we avoid such data, as preswelling may lead to huge variations in the free volume. Even the manufacture of polymeric aerogels, to some extent, falls into this category.

(3.2) **Ordered aggregates** may be formed by some polymers.[344] For example, upon annealing, polyimides form interchain charge-transfer complexes,[345] leading to denser packing and a significant decrease in both free volume and the mean radii of its elements.[346] In a sense, annealed polymers also become somewhat cross-linked and swelling-resistant. A detailed review can be found elsewhere.[347] Polyimides, a crucial class of membrane materials, constitute a considerable share of the current dataset. Membrane film annealing is a standard procedure in most techniques described in the literature, with various protocols and polymer chemical structures causing different degrees of aggregation. For example, compared to pristine Matrimid, annealing it at 350 °C in air leads to a 30% reduction in CO_2 sorption in the low-pressure limit.[348] However, typical annealing protocols usually involve lower temperatures (~200 °C) and less aggressive environments (vacuum, N_2). Any swelling at higher pressures tends to further mitigate the impact of thermal treatment. We roughly estimate that the sorption in an individual polyimide sample falls within a 20% margin relative to the sorption in a sample treated with an "averaged" protocol.

(3.3) **Tacticity**, the relative stereochemistry of the adjoining repeating units within the entire macromolecule, impacts the properties of polymers with chiral repeating units, such as various vinyl polymers. It is not clear whether tacticity is a source of variation, as strictly amorphous polymers of different tacticity frequently differ slightly regarding free volume distribution[349] and sorption capacity.[350] However, data on stereoregular polymers should be approached with caution, as they can easily crystallize during storage or sorption experiments, significantly affecting the measured values.[350]

(3.4) **The chemical composition** of a polymer sample also matters:

- **Chain's tail type** of a polymer might influence its sorption characteristics. Suppose a penetrant gas has a low affinity for the polymer. Then the compatibility of its chain-end functionality with the penetrant could significantly contribute to overall sorption. For instance, in a study by Hallinan *et al.*,[351] in-house synthesized polystyrene with hydroxyl tail groups exhibited a 40–70% increase in water vapor uptake compared to commercial alkyl-terminated polystyrene samples. The importance of chain-tail chemistry increases as the polymer's molecular weight decreases.[352] In the current version of the dataset, the polymers have a high molecular weight; therefore, only a few systems concerning water vapor may exhibit noticeable variance due to the chemistry of the polymers' chain tails.

- **The chemical homogeneity** of some polymers, although formally homopolymers, may be less than ideal. For instance, varying amounts of ethylene's homologs are used as comonomers in the manufacture of certain commercial polyethylenes. Similarly, polymers derived from poly(vinyl alcohol) are not entirely identical, owing to variations in esterification reaction yields or hydrolysis by water vapor during storage.

- **Branching** may be present in polymers that are formally linear. For this reason, HDPE is predominantly used in the dataset.

- **Admixtures**, including plasticizers, stabilizers, casting solvent traces, or components of commercial formulas, are common in polymer samples. For example, Celazole S26 polybenzimidazole is available as a solution in DMAc stabilized by LiCl which cannot be removed along with the solvent under vacuum. The efforts to track these admixtures, account for their effects, purify the samples, or fully specify their origin, composition, and storage conditions vary significantly across the literature.

(3.5) **The aging** of polymer samples is a complex process that may involve physical aging, chemical aging, and absorption aging.[353] Physical aging results from the relaxation of the polymer structure, leading to a reduction in free volume, and it occurs in most glassy polymers. Chemical aging involves interaction with storage atmosphere components, such as oxygen and water vapor. It is more polymer-specific and can range from negligible to very pronounced due to double bonds or ester moieties. An extreme example of aging's impact on sorption is seen in poly(1-trimethylsilyl-1-propyne) (PTMSP), a "high free volume polymer" extensively studied for membrane applications. In a case study by Nagai and Nakagawa[354] propane uptake by PTMSP dropped by 50% after 30 days under vacuum. The extent of this effect varied from 30–50%, depending on conditions, and resulted from a combination of relaxation, oxidation, and even the sorption of vacuum pump oil traces, which are thought to occupy some free volume and enhance polymer matrix relaxation.

The listed items, as one may notice, are partially correlated. For instance, variation in free volume is a key factor underlying most of these items to some extent. However, the reasons for changes in free volume differ. Similarly, aging can be considered a special type of polymer pretreatment, but in the case of aging, the conditions and duration of such "treatment" are typically unknown or imprecise. Additionally, there are other complex issues, such as the interplay between adsorption and absorption in thin films, polymer microporosity, *etc*. Generalizing and formalizing these behaviors into a format suitable for machine learning models presents an opportunity for further research.

5. Conclusion

Convolutional neural networks have been demonstrated to be effective in estimating the solubility of gases, vapors, and supercritical fluids in amorphous polymers, using the electron density distribution of a single polymer repeating unit. In this study, our trial deep learning model converged with an error close to the estimated irreducible error, achieving a remarkable 32% mean absolute error on a set of previously unfamiliar polymers. This included polymers with functional groups not represented in the training set. Notably, the model's predictions were almost independent of the choice of conformers for both the polymer repeating unit and the gas molecule.

However, the model encountered challenges in predicting sorption in polymers with long and flexible substituents, as well as in polymers with polar substituents for water vapor sorption. It also struggled with polymers whose sorption capacity is highly dependent on the polymer's history. For the first two issues, expanding the dataset is likely to offer a solution. For the latter, while the inherent variance in data limits the achievable accuracy of the model, the inclusion of macroscopic features such as the polymer's density could prove beneficial. However, selecting the appropriate macroscopic features and accurately representing the polymer history is a complex task. Among others, it depends on the availability of accurate data and the intended purpose of the model.

The proposed approach is particularly well-suited for exploratory research on hypothetical materials when prior knowledge of their properties is lacking. It relies exclusively on quantum chemical calculations and utilizes a fundamental feature: the electron density distribution. The most exciting aspect of the proposed architecture is its flexibility. Its applicability can be easily extended to, for instance, copolymers, polymers with specific chain ends, or the mixed sorption of multiple gases, simply by adding additional CNN-tails within the twin scheme. Hopefully, in the extended models, one will be able to utilize a pre-trained CNN feature extractor, for example, from the current research, with only the dense network requiring training from scratch.

Another potentially effective strategy is to intentionally narrow the model's scope to a specific class of polymers within a particular pressure and temperature domain. On one hand, such an approach would greatly reduce data variability. On the other hand, it would facilitate the formalization of polymer sample treatment, especially if all the polymers of interest are prepared according to a uniform protocol. For instance, polymers belonging to a certain class may exhibit similar responses to annealing (or any other treatment). Therefore, instead of using the often unavailable sample density, the annealing temperature could be employed as a feature.

Abbreviations

B3LYP	Becke's 3-parameter density functional
CNN	Convolutional neural network
COSMO-RS	Klamt's conductor-like screening model for real solvents
COSMO-SAC	Conductor-like screening model for segment activity coefficient
def2-TZVP	Karlsruhe valence triple-zeta basis set with reparametrized polarization functions
DFT	Density functional theory
DL	Deep learning
DMAc	Dimethylacetamide
DNN	Dense neural network
DSC	Differential scanning calorimetry
ELU	Exponential linear unit
FLOP	Floating-point operations
FTIR	Fourier transform infrared spectroscopy
GFN2-xTB	Grimme's geometry, frequency, noncovalent, extended tight binding method
HDPE	High-density polyethylene
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MC	Monte Carlo method
MD	Molecular dynamics
ML	Machine learning
M_n	Number average molecular weight
MSE	Mean of squares of errors
M_w	Weight average molecular weight
NELF	Non-equilibrium lattice fluid
NET-GP	Non-equilibrium thermodynamics for glassy polymers
NN	Neural network
P	Pressure
P_c	Critical pressure
PC-SAFT	Perturbed chain statistical associating fluid theory
PEG	Poly(ethylene glycol)
PIB	Polyisobutene
PMMA	Poly(methyl methacrylate)
PMP	Poly(4-methyl-2-pentyne)
PPP	Poly(1-phenyl-1-propyne)
$P_{saturated}$	Saturated vapor pressure
PTMSDPA	Poly(1-phenyl-2-[*p*-(trimethylsilyl)phenyl] acetylene)
PTMSP	Poly(1-trimethylsilyl-1-propyne)
PVAc	Polyvinyl acetate
QSAR	Quantitative structure–activity relationships
r^{2}SCAN-3c	Grimme's composite electronic structure method
RI	The resolution of identity approximation
SCF	Supercritical fluid
SL EOS	Sanchez–Lacombe equation of state
SS EOS	Simha–Somcynsky equation of state
SSNMR	Solid-state nuclear magnetic resonance spectroscopy
STP	Standard temperature and pressure
T	Temperature
T_b	Boiling temperature
T_c	Critical temperature

T_g	Glass transition temperature
VOC	Volatile organic compound
X	Crystallinity
XRD	X-ray diffraction analysis
ρ	Density
χ	Molar sorption ratio
ω_s	Solvent weight fraction

Author contributions

Oleg I. Gromov: conceptualization, data curation, software, formal analysis, writing the original draft.

Data availability

The data collection and the code used to automate quantum chemical data generation and formatting and to fit DL models may be accessed at **https://github.com/Shorku/RhNet**.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The author expresses gratitude to Dr E. N. Golubeva and Prof. M. Ya. Melnikov for their support, Mr K. E. Vinogradov for insightful discussions, and Dr I. S. Makarov for reviewing the original draft and providing helpful recommendations. Düziçi'ne ve harika insanlarına özel teşekkürler. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1 R. W. Baker and B. T. Low, *Macromolecules*, 2014, **47**, 6999–7013.
2 M. Tomić, M. Šetka, L. Vojkůvka and S. Vallejos, *Nanomaterials*, 2021, **11**, 552.
3 V. Siracusa, *Int. J. Polym. Sci.*, 2012, **2012**, 1–11.
4 E. Reverchon, R. Adami, S. Cardea and G. Della Porta, *J. Supercrit. Fluids*, 2009, **47**, 484–492.
5 R. K. Kankala, Y. S. Zhang, S.-B. Wang, C.-H. Lee and A.-Z. Chen, *Adv. Healthcare Mater.*, 2017, **6**, 1700433.
6 E. S. Alekseev, A. Y. Alentiev, A. S. Belova, V. I. Bogdan, T. V. Bogdan, A. V. Bystrova, E. R. Gafarova, E. N. Golubeva, E. A. Grebenik, O. I. Gromov, V. A. Davankov, S. G. Zlotin, M. G. Kiselev, A. E. Koklin, Y. N. Kononevich, A. E. Lazhko, V. V. Lunin, S. E. Lyubimov, O. N. Martyanov, I. I. Mishanin, A. M. Muzafarov, N. S. Nesterov, A. Y. Nikolaev, R. D. Oparin, O. O. Parenago, O. P. Parenago, Y. A. Pokusaeva, I. A. Ronova, A. B. Solovieva, M. N. Temnikov, P. S. Timashev, O. V. Turova, E. V. Filatova, A. A. Philippov, A. M. Chibiryaev and A. S. Shalygin, *Russ. Chem. Rev.*, 2020, **89**, 1337–1427.
7 E. Di Maio and E. Kiran, *J. Supercrit. Fluids*, 2018, **134**, 157–166.
8 L. M. Robeson, *J. Membr. Sci.*, 2008, **320**, 390–400.
9 C. Wohlfarth, *CRC Handbook of Thermodynamic Data of Polymer Solutions at Elevated Pressures*, CRC Press, 2005.
10 C. Wohlfarth, *CRC Handbook of Phase Equilibria and Thermodynamic Data of Polymer Solutions at Elevated Pressures*, CRC Press, 2015.
11 R. Paterson, Y. Yampol'skii, P. G. T. Fogg, A. Bokarev, V. Bondar, O. Ilinich and S. Shishatskii, *J. Phys. Chem. Ref. Data*, 1999, **28**, 1255–1450.
12 E. Ricci, M. Minelli and M. G. De Angelis, *Membranes*, 2022, **12**, 857.
13 P. Meares, *J. Am. Chem. Soc.*, 1954, **76**, 3415–3422.
14 P. Meares, *Trans. Faraday Soc.*, 1958, **54**, 40.
15 R. M. Barrer, J. A. Barrie and J. Slater, *J. Polym. Sci.*, 1958, **27**, 177–197.
16 A. S. Michaels, W. R. Vieth and J. A. Barrie, *J. Appl. Phys.*, 1963, **34**, 13–20.
17 W. R. Vieth, P. M. Tam and A. S. Michaels, *J. Colloid Interface Sci.*, 1966, **22**, 360–370.
18 P. J. Flory, *J. Chem. Phys.*, 1942, **10**, 51–61.
19 M. L. Huggins, *J. Phys. Chem.*, 1942, **46**, 151–158.
20 M. Minelli and G. C. Sarti, *Ind. Eng. Chem. Res.*, 2020, **59**, 341–365.
21 F. Doghieri and G. C. Sarti, *Macromolecules*, 1996, **29**, 7885–7896.
22 G. C. Sarti and F. Doghieri, *Chem. Eng. Sci.*, 1998, **53**, 3435–3447.
23 R. H. Lacombe and I. C. Sanchez, *J. Phys. Chem.*, 1976, **80**, 2568–2580.
24 A. I. Papadopoulos, I. Tsivintzelis, P. Linke and P. Seferlis, in *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, Elsevier, 2018.
25 S.-T. Lin and S. I. Sandler, *Ind. Eng. Chem. Res.*, 2002, **41**, 899–913.
26 A. Klamt, V. Jonas, T. Bürger and J. C. W. Lohrenz, *J. Phys. Chem. A*, 1998, **102**, 5074–5085.
27 G. Kupgan, L. J. Abbott, K. E. Hart and C. M. Colina, *Chem. Rev.*, 2018, **118**, 5488–5538.
28 E. N. Muratov, J. Bajorath, R. P. Sheridan, I. V. Tetko, D. Filimonov, V. Poroikov, T. I. Oprea, I. I. Baskin, A. Varnek, A. Roitberg, O. Isayev, S. Curtalolo, D. Fourches, Y. Cohen, A. Aspuru-Guzik, D. A. Winkler, D. Agrafiotis, A. Cherkasov and A. Tropsha, *Chem. Soc. Rev.*, 2020, **49**, 3525–3564.
29 H. Doan Tran, C. Kim, L. Chen, A. Chandrasekaran, R. Batra, S. Venkatram, D. Kamal, J. P. Lightstone, R. Gurnani, P. Shetty, M. Ramprasad, J. Laws, M. Shelton and R. Ramprasad, *J. Appl. Phys.*, 2020, **128**, 171104.

30 H. Ziaee, S. M. Hosseini, A. Sharafpoor, M. Fazavi, M. M. Ghiasi and A. Bahadori, *J. Taiwan Inst. Chem. Eng.*, 2015, **46**, 205–213.
31 H. Karimi, F. Yousefi, E. Ahmadloo and J. Dastranj, *J. Polym. Eng.*, 2014, **34**, 483–488.
32 M. Li, X. Huang, H. Liu, B. Liu, Y. Wu and X. Deng, *J. Appl. Polym. Sci.*, 2013, **129**, 3297–3303.
33 K. Golzar, S. Amjad-Iranagh and H. Modarress, *Measurement*, 2013, **46**, 4206–4225.
34 M. Li, X. Huang, H. Liu, B. Liu, Y. Wu and L. Wang, *RSC Adv.*, 2015, **5**, 45520–45527.
35 A. A. Toropov, A. P. Toropova, S. Begum and P. G. R. Achary, *SAR QSAR Environ. Res.*, 2016, **27**, 293–301.
36 L. Mengshan, W. Wei, C. Bingsheng, W. Yan and H. Xingyuan, *RSC Adv.*, 2017, **7**, 35274–35282.
37 Z. Li, M. Jiang, S. Wang and S. Zhang, *Drug Discovery Today*, 2022, **27**, 103373.
38 Y. Zhao, R. J. Mulder, S. Houshyar and T. C. Le, *Polym. Chem.*, 2023, **14**, 3325–3346.
39 L. Chen, G. Pilania, R. Batra, T. D. Huan, C. Kim, C. Kuenneth and R. Ramprasad, *Mater. Sci. Eng., R*, 2021, **144**, 100595.
40 C. Yan and G. Li, *Adv. Intell. Syst.*, 2023, **5**, 2200243.
41 A. D. Casey, S. F. Son, I. Bilionis and B. C. Barnes, *J. Chem. Inf. Model.*, 2020, **60**, 4457–4473.
42 M. Tsubaki and T. Mizoguchi, *Phys. Rev. Lett.*, 2020, **125**, 206401.
43 A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf and M. Ceriotti, *ACS Cent. Sci.*, 2019, **5**, 57–64.
44 B. Cuevas-Zuviría and L. F. Pacios, *J. Chem. Inf. Model.*, 2020, **60**, 3831–3842.
45 M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller and K. Burke, *Nat. Commun.*, 2020, **11**, 5223.
46 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.
47 B. Liu, M. Wang, H. Foroosh, M. Tappen and M. Penksy, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2015, pp. 806–814.
48 J. P. Boetker, *arXiv*, 2020, preprint, arXiv:2010.13054, **http://arxiv.org/abs/2010.13054**.
49 B. Graham, in *Procedings of the British Machine Vision Conference 2015*, British Machine Vision Association, 2015, pp. 150.1–150.9.
50 Y. Yan, Y. Mao and B. Li, *Sensors*, 2018, **18**, 3337.
51 X. Li, J. Guivant, N. Kwok, Y. Xu, R. Li and H. Wu, *arXiv*, 2019, preprint, arXiv:1901.08373, **http://arxiv.org/abs/1901.08373**.
52 C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.
53 K. Eichkorn, F. Weigend, O. Treutler and R. Ahlrichs, *Theor. Chem. Acc.*, 1997, **97**, 119–124.
54 *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, ed. P. J. Linstrom and W. G. Mallard, National Institute of Standards and Technology, Gaithersburg MD, 2023.
55 Molecular Design LAB. (Dept. of Chemical Engineering), Korea Thermophysical Properties Data Bank (KDB), **https://www.cheric.org/research/kdb/hcprop/cmpsrch.php**.
56 A. Rohatgi, WebPlotDigitizer, **https://automeris.io/WebPlotDigitizer**.
57 G. Landrum.
58 S. Wang, J. Witek, G. A. Landrum and S. Riniker, *J. Chem. Inf. Model.*, 2020, **60**, 2044–2058.
59 T. A. Halgren, *J. Comput. Chem.*, 1999, **20**, 720–729.
60 F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.
61 A. M. Genaev, conformers, **http://limor1.nioch.nsc.ru/quant/program/conformers**, (accessed September 2021).
62 S. Grimme, A. Hansen, S. Ehlert and J.-M. Mewes, *J. Chem. Phys.*, 2021, **154**, 064103.
63 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297.
64 M. Minelli, G. Cocchi, L. Ansaloni, M. G. Baschetti, M. G. De Angelis and F. Doghieri, *Ind. Eng. Chem. Res.*, 2013, **52**, 8936–8945.
65 H. E. Park and J. M. Dealy, *Macromolecules*, 2006, **39**, 5438–5452.
66 E. Kukova, PhD thesis, Ruhr-Universität Bochum, 2003.
67 A. Galia, M. Abduljawad, O. Scialdone and G. Filardo, *AIChE J.*, 2006, **52**, 2243–2253.
68 Y. Sato, M. Wang, S. Takishima, H. Masuoka, T. Watanabe and Y. Fukasawa, *Polym. Eng. Sci.*, 2004, **44**, 2083–2089.
69 J. Chen, T. Liu, L. Zhao and W. Yuan, *Thermochim. Acta*, 2012, **530**, 79–86.
70 T. Guadagno and S. G. Kazarian, *J. Phys. Chem. B*, 2004, **108**, 13995–13999.
71 G. Li, S. N. Leung, M. M. Hasan, J. Wang, C. B. Park and R. Simha, *Fluid Phase Equilib.*, 2008, **266**, 129–142.
72 J.-P. E. Grolier and S. L. Randzio, *J. Chem. Thermodyn.*, 2012, **46**, 42–56.
73 M. Galizia, C. Daniel, G. Fasano, G. Guerra and G. Mensitieri, *Macromolecules*, 2012, **45**, 3604–3615.
74 Y. Sato, T. Takikawa, M. Yamane, S. Takishima and H. Masuoka, *Fluid Phase Equilib.*, 2002, **194–197**, 847–858.
75 Y. Sato, K. Fujiwara, T. Takikawa, S. Sumarno, S. Takishima and H. Masuoka, *Fluid Phase Equilib.*, 1999, **162**, 261–276.
76 Y. Sato, M. Yurugi, K. Fujiwara, S. Takishima and H. Masuoka, *Fluid Phase Equilib.*, 1996, **125**, 129–138.
77 S. Hilic, S. A. E. Boyer, A. A. H. Pádua and J.-P. E. Grolier, *J. Polym. Sci., Part B: Polym. Phys.*, 2001, **39**, 2063–2070.
78 Y. Sato, T. Takikawa, S. Takishima and H. Masuoka, *J. Supercrit. Fluids*, 2001, **19**, 187–198.
79 V. Wiesmet, E. Weidner, S. Behme, G. Sadowski and W. Arlt, *J. Supercrit. Fluids*, 2000, **17**, 1–12.
80 Y. Sato, T. Iketani, S. Takishima and H. Masuoka, *Polym. Eng. Sci.*, 2000, **40**, 1369–1375.
81 S. Areerat, E. Funami, Y. Hayata, D. Nakagawa and M. Ohshima, *Polym. Eng. Sci.*, 2004, **44**, 1915–1924.

82 S. Cotugno, E. Di Maio, G. Mensitieri, S. Iannace, G. W. Roberts, R. G. Carbonell and H. B. Hopfenberg, *Ind. Eng. Chem. Res.*, 2005, **44**, 1795–1803.
83 M. Pantoula, J. von Schnitzler, R. Eggers and C. Panayiotou, *J. Supercrit. Fluids*, 2007, **39**, 426–434.
84 A. de Nicola, A. Correa, G. Milano, P. La Manna, P. Musto, G. Mensitieri and G. Scherillo, *J. Phys. Chem. B*, 2017, **121**, 3162–3176.
85 D. Pierleoni, M. Minelli, G. Scherillo, G. Mensitieri, V. Loianno, F. Bonavolontà and F. Doghieri, *J. Phys. Chem. B*, 2017, **121**, 9969–9981.
86 K.-M. Krüger and G. Sadowski, *Macromolecules*, 2005, **38**, 8408–8417.
87 M. Galizia, M. G. De Angelis, E. Finkelshtein, Y. P. Yampolskii and G. C. Sarti, *J. Membr. Sci.*, 2011, **385–386**, 141–153.
88 J. Dubois, E. Grau, T. Tassaing and M. Dumon, *J. Supercrit. Fluids*, 2018, **131**, 150–156.
89 G. Kravanja, M. K. Hrnčič, M. Škerget and Ž. Knez, *J. Supercrit. Fluids*, 2016, **108**, 45–55.
90 Y. Sun, M. Matsumoto, K. Kitashima, M. Haruki, S. Kihara and S. Takishima, *J. Supercrit. Fluids*, 2014, **95**, 35–43.
91 Y. Sun, M. Matsumoto, M. Haruki, S. Kihara and S. Takishima, *J. Supercrit. Fluids*, 2016, **113**, 144–149.
92 I. Ushiki, S. Hayashi, S. Kihara and S. Takishima, *J. Supercrit. Fluids*, 2019, **152**, 104565.
93 I. S. Liau and M. A. McHugh, in *Supercritical Fluid Technology*, ed. J. M. L. J. M. L. Penninger, M. Radosz, M. A. McHugh and V. J. Krukonis, Elsevier, Amsterdam, 1985, pp. 415–434.
94 A. Garg, E. Gulari and C. W. Manke, *Macromolecules*, 1994, **27**, 5643–5653.
95 W. J. Koros, G. N. Smith and V. Stannett, *J. Appl. Polym. Sci.*, 1981, **26**, 159–170.
96 J. S. Chiou and D. R. Paul, *J. Membr. Sci.*, 1989, **45**, 167–189.
97 N. Platé, A. Bokarev, N. Kaliuzhnyi, E. Litvinova, V. Khotimskii, V. Volkov and Y. Yampol'skii, *J. Membr. Sci.*, 1991, **60**, 13–24.
98 A. C. Puleo, N. Muruganandam and D. R. Paul, *J. Polym. Sci., Part B: Polym. Phys.*, 1989, **27**, 2385–2406.
99 Y. Ichiraku, S. A. Stern and T. Nakagawa, *J. Membr. Sci.*, 1987, **34**, 5–18.
100 Y. Maeda and D. R. Paul, *J. Polym. Sci., Part B: Polym. Phys.*, 1987, **25**, 981–1003.
101 K. Toi, G. Morel and D. R. Paul, *J. Appl. Polym. Sci.*, 1982, **27**, 2997–3005.
102 V. V. Volkov, A. K. Bokarev and S. G. Durgaryan, *Vyoskomol. Soedin, Ser. B*, 1985, **26**, 107.
103 E. M. Davis, M. Minelli, M. G. Baschetti, G. C. Sarti and Y. A. Elabd, *Macromolecules*, 2012, **45**, 7486–7494.
104 E. M. Davis and Y. A. Elabd, *Ind. Eng. Chem. Res.*, 2013, **52**, 12865–12875.
105 M. G. Pastore Carbone, E. Di Maio, S. Iannace and G. Mensitieri, *Polym. Test.*, 2011, **30**, 303–309.
106 E. Markočič, M. Škerget and Ž. Knez, *Ind. Eng. Chem. Res.*, 2013, **52**, 15594–15601.
107 I. Ushiki, H. Kawashima, S. Kihara and S. Takishima, *J. Supercrit. Fluids*, 2022, **181**, 105499.
108 M. G. Pastore Carbone, E. Di Maio, G. Scherillo, G. Mensitieri and S. Iannace, *J. Supercrit. Fluids*, 2012, **67**, 131–138.
109 D. Wang, Z. Cai, X. Huang and L. Wang, *ACS Omega*, 2021, **6**, 1971–1984.
110 N. M. B. Flichy, S. G. Kazarian, C. J. Lawrence and B. J. Briscoe, *J. Phys. Chem. B*, 2002, **106**, 754–759.
111 S. Wang, C. Peng, K. Li, J. Wang, J. Shi and H. Liu, *J. Chem. Ind. Eng.*, 2003, **54**, 141.
112 O. Pfohl, C. Riebesell and R. Dohrn, *Fluid Phase Equilib.*, 2002, **202**, 289–306.
113 R. Pourdarvish, R. P. Danner and J. L. Duda, *J. Appl. Polym. Sci.*, 2008, **108**, 1407–1413.
114 O. Hölck, M. R. Siegert, M. Heuchel and M. Böhning, *Macromolecules*, 2006, **39**, 9590–9604.
115 J.-S. Wang, Y. Kamiya and Y. Naito, *J. Polym. Sci., Part B: Polym. Phys.*, 1998, **36**, 1695–1702.
116 J. S. Chiou, Y. Maeda and D. R. Paul, *J. Appl. Polym. Sci.*, 1987, **33**, 1823–1828.
117 J.-S. Wang and Y. Kamiya, *J. Membr. Sci.*, 1995, **98**, 69–76.
118 A. J. Erb and D. R. Paul, *J. Membr. Sci.*, 1981, **8**, 11–22.
119 H. J. Kim, H. C. Moon, H. Kim, K. Kim, J. K. Kim and J. Cho, *Macromolecules*, 2013, **46**, 493–499.
120 T. C. Merkel, V. Bondar, K. Nagai and B. D. Freeman, *J. Polym. Sci., Part B: Polym. Phys.*, 2000, **38**, 273–296.
121 M. Houben, R. van Geijn, M. van Essen, Z. Borneman and K. Nijmeijer, *J. Membr. Sci.*, 2021, **620**, 118922.
122 J. Guan, Y. Lu, L. Du, C. Liang, J. Wu, D. Li and S. Zhang, *Macromol. Rapid Commun.*, 2022, **43**, 2100796.
123 M. Etxeberria-Benavides, T. Johnson, S. Cao, B. Zornoza, J. Coronas, J. Sanchez-Lainez, A. Sabetghadam, X. Liu, E. Andres-Garcia, F. Kapteijn, J. Gascon and O. David, *Sep. Purif. Technol.*, 2020, **237**, 116347.
124 M. Omidvar, H. Nguyen, L. Huang, C. M. Doherty, A. J. Hill, C. M. Stafford, X. Feng, M. T. Swihart and H. Lin, *ACS Appl. Mater. Interfaces*, 2019, **11**, 47365–47372.
125 S. Singh, A. M. Varghese, K. S. K. Reddy, G. E. Romanos and G. N. Karanikolos, *Ind. Eng. Chem. Res.*, 2021, **60**, 11289–11308.
126 M. Yavari, M. Omidvar and H. Lin, *ACS Appl. Polym. Mater.*, 2019, **1**, 1641–1647.
127 C. R. Maroon, J. Townsend, M. A. Higgins, D. J. Harrigan, B. J. Sundell, J. A. Lawrence, J. T. O'Brien, D. O'Neal, K. D. Vogiatzis and B. K. Long, *J. Membr. Sci.*, 2020, **595**, 117532.
128 C. R. Maroon, J. Townsend, K. R. Gmernicki, D. J. Harrigan, B. J. Sundell, J. A. Lawrence, S. M. Mahurin, K. D. Vogiatzis and B. K. Long, *Macromolecules*, 2019, **52**, 1589–1600.
129 M. Houben, Z. Borneman and K. Nijmeijer, *Sep. Purif. Technol.*, 2021, **255**, 117307.

130 F. Hamidavi, A. Kargari and A. Eliassi, *Sep. Purif. Technol.*, 2021, **279**, 119774.
131 N. M. Larocca and L. A. Pessan, *J. Membr. Sci.*, 2003, **218**, 69–92.
132 J. H. Lee, R. K. M. Chu, R. Li, K. Kwan and C. B. Park, *J. CO_2 Util.*, 2022, **63**, 102131.
133 C. L. Bentley, T. Song, B. J. Pedretti, M. J. Lubben, N. A. Lynd and J. F. Brennecke, *J. Chem. Eng. Data*, 2021, **66**, 2832–2843.
134 S. K. Burgess, O. Karvan, J. R. Johnson, R. M. Kriegel and W. J. Koros, *Polymer*, 2014, **55**, 4748–4756.
135 M. M. López-González, V. Compañ, E. Saiz, E. Riande and J. Guzmán, *J. Membr. Sci.*, 2005, **253**, 175–181.
136 H. Yu, Y. Lei, X. Yu, X. Wang, T. Liu and S. Luo, *J. Polym. Res.*, 2016, **23**, 121.
137 G. Mensitieri, G. Scherillo, P. La Manna and P. Musto, *Membranes*, 2019, **9**, 23.
138 G. Wibawa, R. Hatano, Y. Sato, S. Takishima and H. Masuoka, *J. Chem. Eng. Data*, 2002, **47**, 1022–1029.
139 M. A. Abdulhamid, G. Genduso, Y. Wang, X. Ma and I. Pinnau, *Ind. Eng. Chem. Res.*, 2020, **59**, 5247–5256.
140 M. A. Abdulhamid, G. Genduso, X. Ma and I. Pinnau, *Sep. Purif. Technol.*, 2021, **257**, 117910.
141 A. A. Shamsabadi, A. Kargari, M. B. Babaheidari and S. Laki, *J. Ind. Eng. Chem.*, 2013, **19**, 1680–1688.
142 S. H. Mahmood, M. Keshtkar and C. B. Park, *J. Chem. Thermodyn.*, 2014, **70**, 13–23.
143 G. Wibawa, M. Takahashi, Y. Sato, S. Takishima and H. Masuoka, *J. Chem. Eng. Data*, 2002, **47**, 518–524.
144 G. Wibawa, I. Khoiroh, D. Afrizal and G. Suki, *J. Chem. Eng. Data*, 2010, **55**, 5581–5586.
145 J. P. Stanford, A. L. Maier, L. A. McDonald, P. H. Pfromm and M. E. Rezac, *J. Membr. Sci.*, 2016, **512**, 29–37.
146 X.-K. Li, G.-P. Cao, L.-H. Chen, R.-H. Zhang, H.-L. Liu and Y.-H. Shi, *Langmuir*, 2013, **29**, 14089–14100.
147 H. C. Wong, S. W. Campbell and V. R. Bhethanabotla, *J. Chem. Eng. Data*, 2016, **61**, 3877–3882.
148 H. C. Wong, S. W. Campbell and V. R. Bhethanabotla, *J. Chem. Eng. Data*, 2011, **56**, 4772–4777.
149 H. C. Wong, S. W. Campbell and V. R. Bhethanabotla, *J. Chem. Eng. Data*, 2018, **63**, 2753–2757.
150 K. Kaur, A. R. Iyer, S. W. Campbell and V. R. Bhethanabotla, *J. Chem. Eng. Data*, 2020, **65**, 5046–5054.
151 P. Kundra, S. R. Upreti, A. Lohi and J. Wu, *J. Appl. Polym. Sci.*, 2011, **121**, 2828–2834.
152 J. S. Vrentas, J. L. Duda, H.-C. Ling and A.-C. Hou, *J. Polym. Sci., Polym. Phys. Ed.*, 1985, **23**, 289–304.
153 J. E. Palamara, J. M. Zielinski, M. Hamedi, J. L. Duda and R. P. Danner, *Macromolecules*, 2004, **37**, 6189–6196.
154 J. S. Vrentas, J. L. Duda and A.-C. Hou, *J. Appl. Polym. Sci.*, 1984, **29**, 399–406.
155 H. C. Wong, S. W. Campbell and V. R. Bhethanabotla, *Fluid Phase Equilib.*, 2001, **179**, 181–191.
156 H. C. Wong, S. W. Campbell and V. R. Bhethanabotla, *Fluid Phase Equilib.*, 1997, **139**, 371–389.
157 Y. Iwai and Y. Arai, *J. Chem. Eng. Jpn.*, 1989, **22**, 155–161.
158 P. H. Pfromm and W. J. Koros, *Macromolecules*, 1993, **26**, 6141–6142.
159 T. C. Merkel and L. G. Toy, *Macromolecules*, 2006, **39**, 7591–7600.
160 Y. Liu, Z. Liu, G. Liu, W. Qiu, N. Bhuwania, D. Chinn and W. J. Koros, *J. Membr. Sci.*, 2020, **593**, 117430.
161 H. Sejour, PhD thesis, Georgia Institute of Technology, 2007.
162 A. Venkatasubramanian, PhD thesis, Georgia Institute of Technology, 2013.
163 C. Staudt-Bickel and W. J. Koros, *J. Membr. Sci.*, 2000, **170**, 205–214.
164 S. K. Burgess, D. S. Mikkilineni, D. B. Yu, D. J. Kim, C. R. Mubarak, R. M. Kriegel and W. J. Koros, *Polymer*, 2014, **55**, 6861–6869.
165 A. Kadam, T. Karbowiak, A. Voilley, J.-P. Bellat, O. Vitrac and F. Debeaufort, *J. Polym. Sci., Part B: Polym. Phys.*, 2014, **52**, 1252–1258.
166 J. S. Lee, R. T. Adams, W. Madden and W. J. Koros, *Polymer*, 2009, **50**, 6049–6056.
167 L. M. Costello and W. J. Koros, *J. Polym. Sci., Part B: Polym. Phys.*, 1995, **33**, 135–146.
168 A. Shimazu, T. Miyazaki, M. Maeda and K. Ikeda, *J. Polym. Sci., Part B: Polym. Phys.*, 2000, **38**, 2525–2536.
169 G. Q. Chen, C. A. Scholes, G. G. Qiao and S. E. Kentish, *J. Membr. Sci.*, 2011, **379**, 479–487.
170 C. A. Scholes, W. X. Tao, G. W. Stevens and S. E. Kentish, *J. Appl. Polym. Sci.*, 2010, **117**, 2284–2289.
171 J.-S. Wang, Y. Naito and Y. Kamiya, *J. Polym. Sci., Part B: Polym. Phys.*, 1996, **34**, 2027–2033.
172 K. Tanaka, A. Taguchi, J. Hao, H. Kita and K. Okamoto, *J. Membr. Sci.*, 1996, **121**, 197–207.
173 K. A. Lokhandwala, S. M. Nadakatti and S. A. Stern, *J. Polym. Sci., Part B: Polym. Phys.*, 1995, **33**, 965–975.
174 A. Shimazu, T. Miyazaki, T. Matsushita, M. Maeda and K. Ikeda, *J. Polym. Sci., Part B: Polym. Phys.*, 1999, **37**, 2941–2949.
175 K. Tanaka, H. Kita and K.-I. Okamoto, *J. Polym. Sci., Part B: Polym. Phys.*, 1993, **31**, 1127–1133.
176 A. Shimazu, T. Miyazaki, S. Katayama and Y. Ito, *J. Polym. Sci., Part B: Polym. Phys.*, 2003, **41**, 308–318.
177 O. Rodríguez, F. Fornasiero, A. Arce, C. J. Radke and J. M. Prausnitz, *Polymer*, 2003, **44**, 6323–6333.
178 M. Das and W. J. Koros, *J. Membr. Sci.*, 2010, **365**, 399–408.
179 K.-I. Okamoto, N. Tanihara, H. Watanabe, K. Tanaka, H. Kita, A. Nakamura, Y. Kusuki and K. Nakagawa, *J. Polym. Sci., Part B: Polym. Phys.*, 1992, **30**, 1223–1231.
180 L. Zhang, Y. Xiao, T.-S. Chung and J. Jiang, *Polymer*, 2010, **51**, 4439–4447.
181 B. Schiewe, C. Staudt-Bickel, A. Vuin and G. Wegner, *ChemPhysChem*, 2001, **2**, 211–218.
182 K. Chen, K. Xu, L. Xiang, X. Dong, Y. Han, C. Wang, L.-B. Sun and Y. Pan, *J. Membr. Sci.*, 2018, **563**, 360–370.

183 C. Wu, K. Zhang, H. Wang, Y. Fan, S. Zhang, S. He, F. Wang, Y. Tao, X. Zhao, Y.-B. Zhang, Y. Ma, Y. Lee and T. Li, *J. Am. Chem. Soc.*, 2020, **142**, 18503–18512.
184 S. Japip, K.-S. Liao and T.-S. Chung, *Adv. Mater.*, 2017, **29**, 1603833.
185 S.-X. Cheng, T.-S. Chung, R. Wang and R. H. Vora, *J. Appl. Polym. Sci.*, 2003, **90**, 2187–2193.
186 O. Vopička, K. Pilnáček, P. Číhal and K. Friess, *J. Polym. Sci., Part B: Polym. Phys.*, 2016, **54**, 561–569.
187 O. Vopička, M. G. De Angelis and G. C. Sarti, *J. Membr. Sci.*, 2014, **449**, 97–108.
188 S. Kanehashi, S. Konishi, K. Takeo, K. Owa, H. Kawakita, S. Sato, T. Miyakoshi and K. Nagai, *J. Membr. Sci.*, 2013, **427**, 176–185.
189 L. Liu, PhD thesis, Georgia Institute of Technology, 2015.
190 P. P. Chapala, M. V. Bermeshev, L. E. Starannikova, N. A. Belov, V. E. Ryzhikh, V. P. Shantarovich, V. G. Lakhtin, N. N. Gavrilova, Y. P. Yampolskii and E. S. Finkelshtein, *Macromolecules*, 2015, **48**, 8055–8061.
191 V. Bondar, A. Alentiev, T. Masuda and Y. Yampolskii, *Macromol. Chem. Phys.*, 1997, **198**, 1701–1708.
192 L. Chen, H. Sheth and R. Kim, *Polym. Eng. Sci.*, 2001, **41**, 990–997.
193 R. G. Wissinger and M. E. Paulaitis, *J. Polym. Sci., Part B: Polym. Phys.*, 1987, **25**, 2497–2510.
194 Z. Mogri and D. R. Paul, *Polymer*, 2001, **42**, 7781–7789.
195 V. Carlà, Y. Hussain, C. Grant, G. C. Sarti, R. G. Carbonell and F. Doghieri, *Ind. Eng. Chem. Res.*, 2009, **48**, 3844–3854.
196 H. Guo and V. Kumar, *Polymer*, 2015, **57**, 157–163.
197 M. Pantoula and C. Panayiotou, *J. Supercrit. Fluids*, 2006, **37**, 254–262.
198 H. R. Azimi and M. Rezaei, *J. Chem. Thermodyn.*, 2013, **58**, 279–287.
199 A. Rajendran, B. Bonavoglia, N. Forrer, G. Storti, M. Mazzotti and M. Morbidelli, *Ind. Eng. Chem. Res.*, 2005, **44**, 2549–2560.
200 Z. Lei, H. Ohyabu, Y. Sato, H. Inomata and R. L. Smith, *J. Supercrit. Fluids*, 2007, **40**, 452–461.
201 G. Li, F. Gunkel, J. Wang, C. B. Park and V. Altstädt, *J. Appl. Polym. Sci.*, 2007, **103**, 2945–2953.
202 M. M. Hasan, Y. G. Li, G. Li, C. B. Park and P. Chen, *J. Chem. Eng. Data*, 2010, **55**, 4885–4895.
203 S. Cotugno, E. Di Maio, C. Ciardiello, S. Iannace, G. Mensitieri and L. Nicolais, *Ind. Eng. Chem. Res.*, 2003, **42**, 4398–4405.
204 J. L. Lundberg, E. J. Mooney and C. E. Rogers, *J. Polym. Sci., Part A-2*, 1969, **7**, 947–962.
205 V. Loianno, K. P. Bye, M. Galizia and P. Musto, *J. Polym. Sci.*, 2020, **58**, 2547–2560.
206 J. D. Moon, M. Galizia, H. Borjigin, R. Liu, J. S. Riffle, B. D. Freeman and D. R. Paul, *Macromolecules*, 2018, **51**, 7197–7208.
207 V. P. Talluri, P. Patakova, T. Moucha and O. Vopicka, *Polymer*, 2019, **11**, 1943.
208 H. Feng, T. Hong, S. M. Mahurin, K. D. Vogiatzis, K. R. Gmernicki, B. K. Long, J. W. Mays, A. P. Sokolov, N.-G. Kang and T. Saito, *Polym. Chem.*, 2017, **8**, 3341–3350.
209 S. Ando, A. Yoshida and K. Nagai, *J. Appl. Polym. Sci.*, 2015, **132**, 42208.
210 S. Kanehashi, Y. Koyama, S. Ando, S. Konishi, R. Shindo, S. Miyata, S. Sato, T. Miyakoshi and K. Nagai, *Polym. Int.*, 2014, **63**, 435–444.
211 L. C. Witchey-Lakshmanan, H. B. Hopfenberg and R. T. Chern, *J. Membr. Sci.*, 1990, **48**, 321–331.
212 K.-I. Okamoto, K. Tanaka, H. Kita, A. Nakamura and Y. Kusuki, *J. Polym. Sci., Part B: Polym. Phys.*, 1989, **27**, 2621–2635.
213 A. X. Wu, J. A. Drayton, K. Mizrahi Rodriguez, F. M. Benedetti, Q. Qian, S. Lin and Z. P. Smith, *Macromolecules*, 2021, **54**, 22–34.
214 H. D. Kamaruddin and W. J. Koros, *J. Polym. Sci., Part B: Polym. Phys.*, 2000, **38**, 2254–2267.
215 X. Duthie, S. Kentish, C. Powell, K. Nagai, G. Qiao and G. Stevens, *J. Membr. Sci.*, 2007, **294**, 40–49.
216 I. Ushiki, S. Ota, S. Kihara and S. Takishima, *J. Supercrit. Fluids*, 2023, **194**, 105836.
217 H. Azimi, D. Jahani and M. Nofar, *J. Chem. Eng. Data*, 2020, **65**, 4596–4604.
218 D. Borrmann, A. Danzer and G. Sadowski, *Membranes*, 2022, **12**, 434.
219 J. M. Zielinski, B. T. Carvill, S. A. Gardner, M. F. Kimak, R. Horvath and J. E. Rovira, *Ind. Eng. Chem. Res.*, 2001, **40**, 2990–2994.
220 E. Sada, H. Kumazawa and P. Xu, *J. Appl. Polym. Sci.*, 1988, **35**, 1497–1509.
221 J. Huang, R. J. Cranford, T. Matsuura and C. Roy, *J. Appl. Polym. Sci.*, 2003, **87**, 2306–2317.
222 H. Tang, W. Sun, H. Plancher, M. Radosz and Y. Shen, *Chem. Commun.*, 2005, 3325.
223 J. Tang, H. Tang, W. Sun, M. Radosz and Y. Shen, *Polymer*, 2005, **46**, 12460–12467.
224 J. Tang, H. Tang, W. Sun, M. Radosz and Y. Shen, *J. Polym. Sci., Part A: Polym. Chem.*, 2005, **43**, 5477–5489.
225 G. Yu, Q. Li, N. Li, Z. Man, C. Pu, C. Asumana and X. Chen, *Polym. Eng. Sci.*, 2014, **54**, 59–63.
226 A. Mehrdad and N. Noorani, *J. Mol. Liq.*, 2019, **291**, 111288.
227 N. Noorani and A. Mehrdad, *J. Polym. Res.*, 2021, **28**, 346.
228 S. Prager, E. Bagley and F. A. Long, *J. Am. Chem. Soc.*, 1953, **75**, 2742–2745.
229 D. Kalmykov, A. Balynin, A. Yushkin, E. Grushevenko, S. Sokolov, A. Malakhov, A. Volkov and S. Bazhenov, *Membranes*, 2022, **12**, 1160.
230 C. Panayiotou and J. H. Vera, *Polym. J.*, 1984, **16**, 89–102.
231 A. Arce, F. Fornasiero, O. Rodríguez, C. J. Radke and J. M. Prausnitz, *Phys. Chem. Chem. Phys.*, 2004, **6**, 103–108.
232 G. Giacobazzi, C. Gioia, M. Vannini, P. Marchese, V. Guillard, H. Angellier-Coussy and A. Celli, *Polymer*, 2021, **13**, 524.

233 J. O. Tanbonliong and J. M. Prausnitz, *Polymer*, 1997, **38**, 5775–5783.
234 J. G. Lieu, J. M. Prausnitz and M. Gauthier, *Polymer*, 2000, **41**, 219–224.
235 J. G. Lieu and J. M. Prausnitz, *Polymer*, 1999, **40**, 5865–5871.
236 C. Mio and J. Prausnitz, *Polymer*, 1998, **39**, 6401–6405.
237 K. Wang, Y. Hu and D. T. Wu, *J. Chem. Eng. Data*, 1994, **39**, 916–919.
238 R. A. G. Sé and M. Aznar, *Polymer*, 2007, **48**, 5646–5652.
239 F. Fornasiero, M. Halim and J. M. Prausnitz, *Macromolecules*, 2000, **33**, 8435–8442.
240 Y. Chang, J. Noriyan, D. R. Lloyd and J. W. Barlow, *Polym. Eng. Sci.*, 1987, **27**, 693–702.
241 E. M. Davis and Y. A. Elabd, *J. Phys. Chem. B*, 2013, **117**, 10629–10640.
242 F. Dubelley, E. Planes, C. Bas, E. Pons, B. Yrieix and L. Flandin, *J. Phys. Chem. B*, 2017, **121**, 1953–1962.
243 G. Scherillo, P. La Manna, P. Musto and G. Mensitieri, *Sci. China: Phys., Mech. Astron.*, 2020, **63**, 247012.
244 G. Scherillo, G. Mensitieri, A. Baldanza, V. Loianno, P. Musto, M. Pannico, A. Correa, A. De Nicola and G. Milano, *Macromolecules*, 2022, **55**, 10773–10787.
245 V. Loianno, A. Baldanza, G. Scherillo, P. Musto and G. Mensitieri, *Polymer*, 2023, **15**, 1144.
246 C. A. Scholes, J. Jin, G. W. Stevens and S. E. Kentish, *J. Polym. Sci., Part B: Polym. Phys.*, 2015, **53**, 719–728.
247 K. P. Bye, V. Loianno, T. N. Pham, R. Liu, J. S. Riffle and M. Galizia, *J. Membr. Sci.*, 2019, **580**, 235–247.
248 M. M. Merrick, PhD thesis, The University of Texas at Austin, 2020.
249 A. Fernández-Barquín, R. Rea, D. Venturi, M. Giacinti-Baschetti, M. G. De Angelis, C. Casado-Coterillo and Á. Irabien, *RSC Adv.*, 2018, **8**, 3536–3546.
250 C. Tsioptsias and C. Panayiotou, *Thermochim. Acta*, 2011, **521**, 98–106.
251 J. Chmelar, T. Gregor, H. Hajova, A. Nistor and J. Kosek, *Polymer*, 2011, **52**, 3082–3091.
252 H. Hajova, J. Chmelar, A. Nistor, T. Gregor and J. Kosek, *J. Chem. Eng. Data*, 2013, **58**, 851–865.
253 H. B. Hopfenberg, R. H. Holley and V. Stannett, *Polym. Eng. Sci.*, 1969, **9**, 242–249.
254 B. R. Baird, H. B. Hopfenberg and V. T. Stannett, *Polym. Eng. Sci.*, 1971, **11**, 274–283.
255 P. Chandra and W. J. Koros, *Polymer*, 2009, **50**, 236–244.
256 L. C. Witchey-Lakshmanan, H. B. Hopfenberg and R. T. Chern, *J. Polym. Sci., Part B: Polym. Phys.*, 1993, **31**, 1545–1553.
257 S. A. Stern, U. M. Vakil and G. R. Mauze, *J. Polym. Sci., Part B: Polym. Phys.*, 1989, **27**, 405–429.
258 Z. Miyagi and K. Tanaka, *Kolloid Z. Z. Polym.*, 1973, **251**, 739–744.
259 Z. Miyagi and K. Tanaka, *Polymer*, 1975, **16**, 441–444.
260 O. Esekhile, W. Qiu and W. J. Koros, *J. Polym. Sci., Part B: Polym. Phys.*, 2011, **49**, 1605–1620.
261 Z. Miyagi and K. Tanaka, *Colloid Polym. Sci.*, 1979, **257**, 259–265.
262 Y. Kamiya, K. Mizoguchi, Y. Naito and T. Hirose, *J. Polym. Sci., Part B: Polym. Phys.*, 1986, **24**, 535–547.
263 Y. Kamiya, T. Hirose, K. Mizoguchi and Y. Naito, *J. Polym. Sci., Part B: Polym. Phys.*, 1986, **24**, 1525–1539.
264 S. V. Dixon-Garrett, K. Nagai and B. D. Freeman, *J. Polym. Sci., Part B: Polym. Phys.*, 2000, **38**, 1078–1089.
265 J. S. McHattie, W. J. Koros and D. R. Paul, *Polymer*, 1991, **32**, 840–850.
266 Y. P. Yampol'skii, V. V. Volkov, N. E. Kaliuzhnyi and S. G. Durgaryan, *Vysokomol. Soedin., Ser. A*, 1984, **26**, 1640.
267 V. V. Volkov, A. K. Bokarev and S. G. Durgaryan, *Vysokomol. Soedin., Ser. A*, 1984, **26**, 1294.
268 C. L. Aitken, W. J. Koros and D. R. Paul, *Macromolecules*, 1992, **25**, 3651–3658.
269 S. Netramai, M. Rubino, R. Auras and B. A. Annous, *J. Appl. Polym. Sci.*, 2009, **114**, 2929–2936.
270 S. Wachi, H. Morikawa and H. Inoue, *AIChE J.*, 1988, **34**, 1683–1690.
271 P. L. Durrill and R. G. Griskey, *AIChE J.*, 1966, **12**, 1147–1151.
272 V. V. Volkov, S. G. Durgaryan, E. G. Novitskii and N. S. Nametkin, *Vysokomol. Soedin., Ser. A*, 1979, **21**, 927.
273 E. S. Sanders and W. J. Koros, *J. Polym. Sci., Part B: Polym. Phys.*, 1986, **24**, 175–188.
274 A. Naderi, A. Asadi Tashvigh and T.-S. Chung, *J. Membr. Sci.*, 2019, **572**, 343–349.
275 A. Naderi, W. F. Yong, Y. Xiao, T.-S. Chung, M. Weber and C. Maletzko, *Polymer*, 2018, **135**, 76–84.
276 K. Nakanishi, H. Odani, M. Kurata, T. Masuda and T. Higashimura, *Polym. J.*, 1987, **19**, 293–296.
277 A. Morisato, B. D. Freeman, I. Pinnau and C. G. Casillas, *J. Polym. Sci., Part B: Polym. Phys.*, 1996, **34**, 1925–1934.
278 L. G. Toy, K. Nagai, B. D. Freeman, I. Pinnau, Z. He, T. Masuda, M. Teraguchi and Y. P. Yampolskii, *Macromolecules*, 2000, **33**, 2516–2524.
279 F. Doghieri and G. C. Sarti, *J. Polym. Sci., Part B: Polym. Phys.*, 1997, **35**, 2245–2258.
280 J. P. Stanford, P. H. Pfromm and M. E. Rezac, *J. Appl. Polym. Sci.*, 2017, **134**, 44771.
281 K. Nagai, S. Kanehashi, S. Tabei and T. Nakagawa, *J. Membr. Sci.*, 2005, **251**, 101–110.
282 S. M. Nadakatti, J. H. Kim and S. A. Stern, *J. Membr. Sci.*, 1995, **108**, 279–291.
283 A. V. Volkov, E. V. Fedorov, A. O. Malakhov and V. V. Volkov, *Polym. Sci., Ser. B*, 2002, **44**, 158–162.
284 H. Okuno, K. Renzo and T. Uragami, *J. Membr. Sci.*, 1995, **103**, 31–38.
285 Y. Kamiya, T. Hirose, K. Mizoguchi and K. Terada, *J. Polym. Sci., Part B: Polym. Phys.*, 1988, **26**, 1409–1424.
286 J. S. Chiou and D. R. Paul, *J. Appl. Polym. Sci.*, 1986, **32**, 4793–4814.
287 J. L. Williams, H. B. Hopfenberg and V. Stannett, *J. Macromol. Sci., Part B: Phys.*, 1969, **3**, 711–725.

288 K. Nagai, L. G. Toy, B. D. Freeman, M. Teraguchi, G. Kwak, T. Masuda and I. Pinnau, *J. Polym. Sci., Part B: Polym. Phys.*, 2002, **40**, 2228–2236.
289 D. C. Bonner and J. M. Prausnitz, *J. Polym. Sci., Polym. Phys. Ed.*, 1974, **12**, 51–73.
290 E. C. Baughan, *Trans. Faraday Soc.*, 1948, **44**, 495.
291 C. E. H. Bawn, R. F. J. Freeman and A. R. Kamaliddin, *Trans. Faraday Soc.*, 1950, **46**, 677–684.
292 B. E. Eichinger and P. J. Flory, *Trans. Faraday Soc.*, 1968, **64**, 2035–2052.
293 C. E. H. Bawn and R. D. Patel, *Trans. Faraday Soc.*, 1956, **52**, 1664–1668.
294 B. E. Eichinger and P. J. Flory, *Trans. Faraday Soc.*, 1968, **64**, 2053–2060.
295 B. E. Eichinger and P. J. Flory, *Trans. Faraday Soc.*, 1968, **64**, 2061–2065.
296 C. Booth and C. J. Devoy, *Polymer*, 1971, **12**, 320–326.
297 A. Nakajima, H. Yamakawa and I. Sakurada, *J. Polym. Sci.*, 1959, **35**, 489–495.
298 C. E. H. Bawn and M. A. Wajid, *Trans. Faraday Soc.*, 1956, **52**, 1658–1664.
299 B. E. Eichinger and P. J. Flory, *Trans. Faraday Soc.*, 1968, **64**, 2066–2072.
300 C. Baker, W. Brown, G. Gee, J. Rowlinson, D. Stubley and R. Yeadon, *Polymer*, 1962, **3**, 215–230.
301 S. Saeki and D. C. Bonner, *Polymer*, 1978, **19**, 319–324.
302 W. C. Dale and C. E. Rogers, *J. Appl. Polym. Sci.*, 1972, **16**, 21–36.
303 N. S. Oliveira, J. Oliveira, T. Gomes, A. Ferreira, J. Dorgan and I. M. Marrucho, *Fluid Phase Equilib.*, 2004, **222–223**, 317–324.
304 J. Trifol, D. Plackett, P. Szabo, A. E. Daugaard and M. Giacinti Baschetti, *ACS Omega*, 2020, **5**, 15362–15369.
305 J. L. Garcia-Fierro and J. V. Aleman, *Eur. Polym. J.*, 1985, **21**, 753–756.
306 A. G. Day, *Trans. Faraday Soc.*, 1963, **59**, 1218.
307 S. U. Hong, T. A. Barbari and J. M. Sloan, *J. Polym. Sci., Part B: Polym. Phys.*, 1998, **36**, 337–344.
308 H. Fujita, A. Kishimoto and K. Matsumoto, *Trans. Faraday Soc.*, 1960, **56**, 424.
309 H. Odani, M. Uchikura and M. Kurata, *Bull. Inst. Chem. Res., Kyoto Univ.*, 1984, **62**, 188–197.
310 J. S. Chiou, J. W. Barlow and D. R. Paul, *J. Appl. Polym. Sci.*, 1985, **30**, 1173–1186.
311 E. M. Nunez, A. S. Myerson and T. K. Kwei, *Polym. Eng. Sci.*, 1991, **31**, 1172–1175.
312 V. Stannett and J. L. Williams, *J. Polym. Sci., Part C: Polym. Symp.*, 1965, **10**, 45–59.
313 Y. Kamiya, D. Bourbon, K. Mizoguchi and Y. Naito, *Polym. J.*, 1992, **24**, 443–449.
314 V. Vittoria and A. R. Filho, *J. Appl. Polym. Sci.*, 1992, **44**, 2127–2138.
315 T. T. Moore and W. J. Koros, *J. Appl. Polym. Sci.*, 2007, **104**, 4053–4059.
316 T.-S. Chung, S. S. Chan, R. Wang, Z. Lu and C. He, *J. Membr. Sci.*, 2003, **211**, 91–99.
317 L. Shao, J. Samseth and M.-B. Hägg, *Int. J. Greenh. Gas Control*, 2008, **2**, 492–501.
318 T. C. Merkel, B. D. Freeman, R. J. Spontak, Z. He, I. Pinnau, P. Meakin and A. J. Hill, *Chem. Mater.*, 2003, **15**, 109–123.
319 O. I. Gromov, RhNet. **https://github.com/Shorku/rhnet**, **https://github.com/Shorku/rhnet**.
320 Y. Taigman, M. Yang, M. Ranzato and L. Wolf, in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2014, pp. 1701–1708.
321 C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2015, pp. 1–9.
322 G. Huang, Z. Liu, G. Pleiss, L. van der Maaten and K. Q. Weinberger, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022, **44**, 8704–8716.
323 M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, *arXiv*, 2016, preprint, arXiv:1603.04467, **http://arxiv.org/abs/1603.04467**.
324 F. Chollet, *Keras*, **https://keras.io**, 2015.
325 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Pietro Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko and Y. Vázquez-Baeza, *Nat. Methods*, 2020, **17**, 261–272.

326 S. Skjold-Jorgensen, B. Kolbe, J. Gmehling and P. Rasmussen, *Ind. Eng. Chem. Process Des. Dev.*, 1979, **18**, 714–722.
327 J. Gmehling, P. Rasmussen and A. Fredenslund, *Ind. Eng. Chem. Process Des. Dev.*, 1982, **21**, 118–127.
328 Y. Sato, T. Takikawa, A. Sorakubo, S. Takishima, H. Masuoka and M. Imaizumi, *Ind. Eng. Chem. Res.*, 2000, **39**, 4813–4819.
329 S. Kanehashi and K. Nagai, *J. Membr. Sci.*, 2005, **253**, 117–138.
330 J. Zhao, X. Li, X. Wang, Q. Zhang, Q. Yang, H. Yin, S. Zhang and C. Wu, *Int. J. Hydrogen Energy*, 2023, **48**, 19619–19632.
331 T. Prodpran, S. Shenogin and S. Nazarenko, *Polymer*, 2002, **43**, 2295–2309.
332 A. Puleo, D. Paul and P. Wong, *Polymer*, 1989, **30**, 1357–1366.
333 M. Sivakumar, T. Suzuki, Y. Yamamoto, K. P. O. Mahesh, H. Yoshimizu and Y. Tsujita, *J. Membr. Sci.*, 2004, **238**, 75–81.
334 O. Atiq, E. Ricci, M. G. Baschetti and M. G. De Angelis, *Fluid Phase Equilib.*, 2022, **556**, 113412.
335 S. Venkatram, J. McCollum, N. Stingelin and B. Brettmann, *Polym. Int.*, 2023, **72**, 855–860.
336 T. B. Martin and D. J. Audus, *ACS Polym. Au*, 2023, **3**, 239–258.
337 W. J. Box, M. T. Webb and M. Galizia, *Ind. Eng. Chem. Res.*, 2022, **61**, 9856–9868.
338 J. K. Lee, S. X. Yao, G. Li, M. B. G. Jun and P. C. Lee, *Polym. Rev.*, 2017, **57**, 695–747.
339 E. Kiran, J. A. Sarver and J. C. Hassler, *J. Supercrit. Fluids*, 2022, **185**, 105378.
340 G. Li, H. Li, J. Wang and C. B. Park, *Cell. Polym.*, 2006, **25**, 237–248.
341 A. V. Ewing and S. G. Kazarian, *J. Supercrit. Fluids*, 2018, **134**, 88–95.
342 J. M. H. Fechter, H. B. Hopfenberg and W. J. Koros, *Polym. Eng. Sci.*, 1981, **21**, 925–929.
343 A. H. Chan and D. R. Paul, *Polym. Eng. Sci.*, 1980, **20**, 87–94.
344 C. A. O'Mahoney, D. J. Williams, H. M. Colquhoun, R. Mayo, S. M. Young, A. Askari, J. Kendrick and E. Robson, *Macromolecules*, 1991, **24**, 6527–6530.
345 F. J. Dinan, W. T. Schwartz, R. A. Wolfe, D. S. Hojnicki, T. St. Clair and J. R. Pratt, *J. Polym. Sci., Part A: Polym. Chem.*, 1992, **30**, 111–118.
346 X. Duthie, S. Kentish, S. J. Pas, A. J. Hill, C. Powell, K. Nagai, G. Stevens and G. Qiao, *J. Polym. Sci., Part B: Polym. Phys.*, 2008, **46**, 1879–1890.
347 Y. Zhuang, J. G. Seong and Y. M. Lee, *Prog. Polym. Sci.*, 2019, **92**, 35–88.
348 M. Houben, J. Kloos, M. van Essen, K. Nijmeijer and Z. Borneman, *J. Membr. Sci.*, 2022, **647**, 120292.
349 R. M. Dammert, S. L. Maunu, F. H. J. Maurer, I. M. Neelov, S. Niemelä, F. Sundholm and C. Wästlund, *Macromolecules*, 1999, **32**, 1930–1938.
350 V. Vittoria and A. R. Filho, *J. Appl. Polym. Sci.*, 1993, **49**, 247–256.
351 D. T. Hallinan, M. Minelli, O. Oparaji, A. Sardano, O. Iyiola, A. R. Garcia and D. J. Burnett, *Membranes*, 2022, **12**, 1059.
352 E. Weidner and V. Wiesmet, in *Thermodynamic Properties of Complex Fluid Mixtures*, ed. G. Maurer, Wiley-VCH, 2004, p. 511.
353 L. Strannikova, V. Khodzhaeva and Y. Yampolskii, *J. Membr. Sci.*, 2004, **244**, 183–191.
354 K. Nagai and T. Nakagawa, *J. Membr. Sci.*, 1995, **105**, 261–272.

