Adv Syst Sci Appl 2020; **03**; 91-104

Classifying Amorphous Polymers for Membrane Technology Basing on Accessible Surface Area of Their Conformations

Oleg Miloserdov[1*]

[1)] *Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*

E-mail: oleg_milos@mail.ru

Abstract: Almost 400 amorphous polymer materials used in membrane gas separation technology are clustered on the basis of the shape of their polymer chains conformations. Obtained clusters, which rely solely on the geometry of polymer chains and not on the chemical class (polyamides, polyacetylenes, etc.), are shown to discriminate polymers with respect to their transport properties, in particular, the coefficient of diffusion. The method proposed consists of several steps. Firstly, realistic conformations of large polymer macromolecules are constructed using the program code developed in the RDKit environment for Python. Then, polymer conformations are characterized by a curve that relates the “accessible surface area” (i.e., the contact surface between the spherical model of a macromolecule and a spherical “probe”) to the radius of this probe, and also seven similar curves, which relate the polarized (neutral, positively or negatively charged, etc.) accessible surface area to the radius of the spherical probe that represents the variety of penetrant gases. An improved algorithm for surface area calculation maps out the outer surface of the macromolecule to eliminate its influence. The curves are averaged between ten polymer conformations to obtain more robust figures. Finally, agglomerative clustering is used to separate different polymers in the space of these curves that align their accessible-surface-area-related quantities against the probe radius. The proposed classification of polymers can be used to develop more precise predictive models of polymers’ transport properties for the theory-guided and computer-aided materials design.

Keywords: machine learning, agglomerative clustering, QSPR, molecular modeling, polymer membranes, gas separation

1. INTRODUCTION

Membrane technologies of gas separation are extensively used all over the world in Hydrogen and Oxygen production, natural gas purification and carbon dioxide separation. Despite the great variety of materials used in liquid, metal, ceramic and polymer membranes, there is still a high need for a scientifically-based search for new materials, especially for polymer membranes. The development of mathematical models to predict properties of polymers from their structure allows saving financial and time resources during the development of new membrane materials.

Unfortunately, the mathematical models for transport properties of polymeric materials so far have either insufficient predictive power or are limited to specific polymers or their classes. Primarily, this is due to the fact that transport properties of polymers used for membrane gas separation are determined by the geometry of their polymer chains, characterized by a wide variety of chemical structures and methods of their organization (spatial isomerism, statistical sequence of polymer units, etc.).

Several approaches are known from the literature to predict transport properties of polymers. Those most popular and well-founded are based on computer simulation of atoms and their interactions [19,8]. First of all, they include molecular dynamics (MD) and the Grand Canonical Monte Carlo Method (GCMC) [13] that are used to predict solubility and

---

[*] Corresponding author: oleg_milos@mail.ru

diffusion coefficients. While MD is used to predict diffusion, GCMC models have shown their effectiveness in predicting solubility coefficients. One of the common shortcomings of both MD and GCMC is their high computational complexity. They require a lot of computing resources and machine time. Therefore, these methods seem to be too time-consuming for mass prediction of the properties (screening) of hypothetical (previously not synthesized) polymeric materials.

At the same time, there is an alternative approach to the prediction of sorption properties of polymers and the prediction of their transport parameters. This is a typical task for QSPR methods [17,21]. Their goal is to establish statistically significant correlations between the numerical values of the physicochemical properties of materials and molecular descriptors.

Earlier, in [11], the method was proposed and then improved in [10] for predicting the sorption parameters of polymers (gas solubility coefficient at infinite dilution S and the solubility constant of Henry's law k_D in the model of double sorption) for a number of light gases in glassy polymers. The approach [11,10] is based on calculating the dependence of a number of geometric indices on the effective penetrant radius for conformations of a polymer chain segment (200-600 atoms length), obtained from computer modeling of a polymer chain segment.

In this article this approach to molecular modeling and data analysis is further developed and applied to building a classification of glassy polymers used in membrane technology. The isolated clusters turn out to be closely related to the physicochemical properties of substances that are important for the membrane technology. In the future, it can be used to construct refined regressions that can predict the main transport properties of polymeric materials — solubility, diffusion, and permeability coefficients.

2. MOLECULAR MODELING

2.1. Molecular Mechanics Modeling

Problems of quantitative property forecasting from statistics and those of massive polymer classification impose some requirements on the underlying approach to molecular mechanicsmodeling (hereinafter MMM).

1. Conformations (macromolecule atoms' positions on a scene) obtained from MMM should be as realistic as possible (conform to real positions of atoms in a polymer membrane).
2. Manual tuning of the MMM algorithm should be avoided, and the algorithm should be as automated as possible.
3. The MMM algorithm should be able to parallelize on a server and have an acceptable calculation time for one polymer.
4. The method should ensure the stability of the results and their reproducibility.
5. The method should be applicable to specific polymers used in membrane gas separation.
6. The method should use the free software.

In this paper, we further improve the MMM method from [11,10], which is based on the modeling of conformations of the polymer macromolecule of about several hundred atoms size. In [11], for each polymer, an oligomer (200–600 atoms length) consisting of several polymer monomer units, was created in the InstantJChem molecular simulation environment [4]. For this chain, using Conformer Plugin [6] of the InstantJChem ChemAxon package, several conformations of a short chain segment of the polymer under consideration (typically including several monomer units and from 200 to 600 atoms) are generated by relaxation in the empirical Dreiding field from random initial positions of atoms. Further calculation of the indices necessary for constructing the regression model was also performed using InstantJChem tools. Despite the fact that, using the obtained conformations, one of the transport parameters of polymer membranes, S solubility, was successfully predicted, this approach did not meet many criteria, listed above. The most important drawback is the unrealistic nature of the many conformations obtained, as well as the difficulty in expanding this solution to server platforms.

In [10], the problem of the unreality of the obtained conformations was solved by performing their modeling in a program in the PerkinElmer Chem3D package [5] using a similar technique. To assign a random initial position of atoms to the simulated polymer chain, molecular mechanics simulation was performed at a temperature of 300 K to 3000 K for 1000 iterations, after which the resulting structure was optimized by free energy in the empirical field MM2. This approach showed good results in terms of realistic conformations, but made it impossible to automate the MMM process and did not allow the efficient use of multiprocessor systems.

During the search, many different MMM solutions were tested, for example, LAMMPS [12], OpenMM [7], Hoomd-Blue [3,9] and others. Some solutions did not allow automation of the process (LAMMPS), the others (e.g., OpenMM) were developed for biopolymers and had no empirical force fields relevant for glassy polymers used in membrane gas separation.

RDKit package [16] - an open-source toolkit – is widely used by the scientific community to solve various problems in chemoinformatics and machine learning. The main data structures and RDKit algorithms are written in C++, which ensures high performance. RDKit also has shells in Python, Java, and C #, which makes it easy to use.

To arrange the molecule in 3D space in the RDKit environment, the flexible EmbedMolecule procedure is used [15]. The initial coordinates of atoms positions in molecule can be determined both through the eigenvalues of the distance matrix, and through a random positioning of atoms, and the value of the random number generator can be fixed, which will later make it possible to obtain the same molecule. Then MMM is performed in an empirical force field UFF (Universal Force Field [14]). It is important that RDKit tools allow to fix the part of a molecule during MMM.

The polymer chain is built by a simple algorithm, which uses the EmbedMolecule function of RDKit and roughly mimics the real process of polymerization. At the beginning, the monomer unit of the polymer is located in the space, then a similarly processed unit is added to it, but with other random coordinates. All atomic coordinates, except the atomic coordinates of the last two added monomer units, are fixed, and two free units undergo energy minimization in the UFF field, which simulates the process of sequential polymerization of a macromolecule. In this case, the algorithm takes into account the experimental values of the torsion angles and uses the base of universal rules for the mutual arrangement of atoms, for example, the mandatory presence in the same plane of the atoms of the benzene ring.

In less than half of the samples, due to the peculiarities in their structure, attempts to construct a macromolecule in this way failed after several iterations. In this case, the procedure of placing the atoms of a molecule in space was performed with the use of random initial coordinates for the entire macromolecule, and not its individual monomer units. Examples of the obtained conformations are presented in Fig. 2.1.

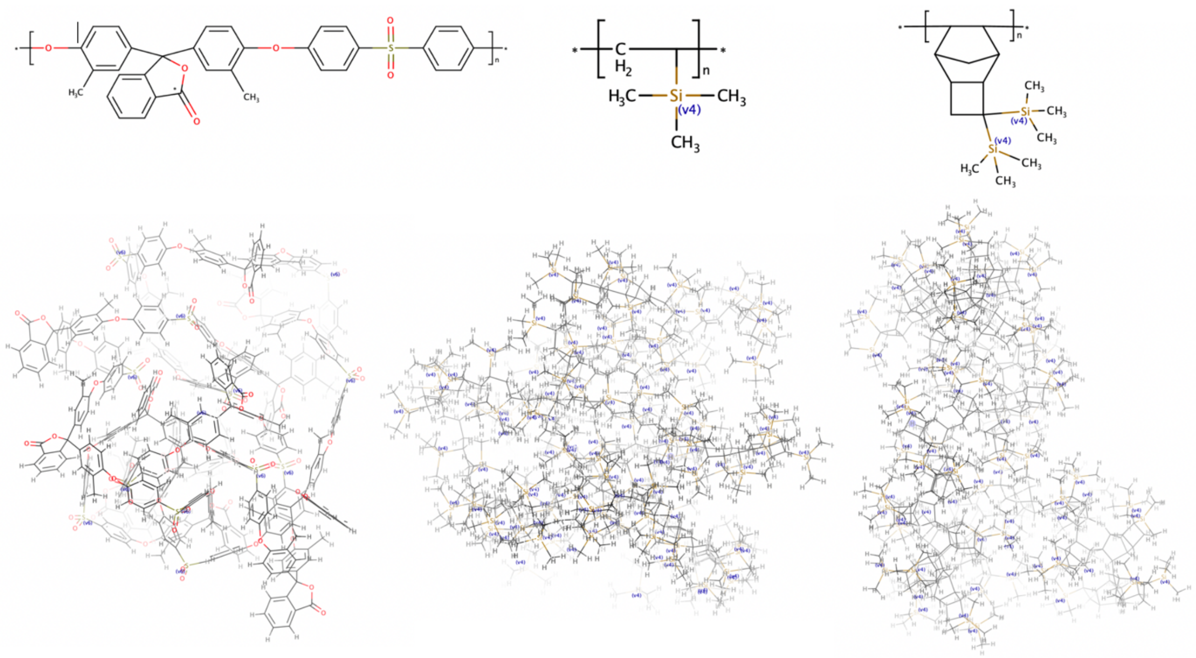

Fig. 2.1. Examples of three polymer repeating units are shown above, and the corresponding macromolecules of about 700-1200 atoms size obtained with MMM are shown below

2.2. Calculation of Geometric Indices

As in [10], the curves that represent the value of various accessible-surface-area-based geometric indices of polymers' conformations as functions of the radius of a spherical "probe" are used to characterize polymers, since the main hypothesis of this method is that transport properties of a polymer (e.g., solubility coefficient and the similar measures) to some extent depend on the characteristics of the contact surface between the polymer macromolecule and the penetrant gas molecule.

The basic and the simplest index of this family is the accessible surface area (ASA). Its calculation is illustrated by Fig. 2.2. ASA is the surface area circumscribed by the center of the spherical "probe" of a given radius in all possible positions of its contact with the van der Waals surface of this molecule.

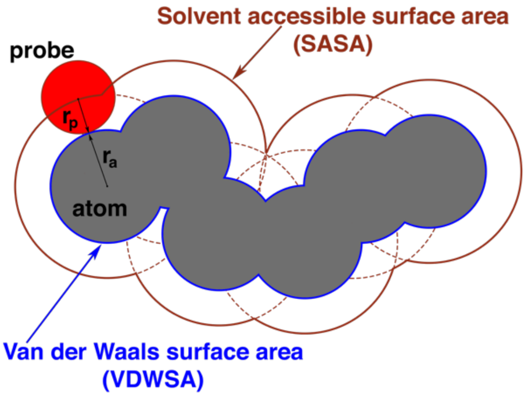

Fig. 2.2. Examples of calculating the accessible surface area of a molecule

The whole family of the employed indices is presented in Table 2.1.

Table 2.1. Geometric indices

Index	Index Description
ASA	Standardized (per cm^3 of the sample) accessible surface area (Å^2·mol/cm^3)
ASA^+	Accessible surface area where contact occurs at a point on the surface with a partial positive ($q > 0$) charge* (Å^2·mol/cm^3)
ASA^-	Accessible surface area where contact occurs at a point on the surface with a partial negative ($q < 0$) charge* (Å^2·mol/cm^3)
ASA_H	Accessible area of hydrophobic (with low, $\|q\| < 0.125$, level of partial charge*) surface (Å^2·mol/cm^3)
ASA_P	Accessible area of polarized (with high, $\|q\| \geq 0.125$, level of partial charge*) surface (Å^2·mol/cm^3)
$DPSA_3$	$DPSA_3 = \sum_i asa_i \cdot q_i$, where asa_i – is the contribution of the i-th atom to the specific accessible surface area of the molecule and q_i is the partial charge* of i-th atom (Å^2 **e** моль/см^3) *
$PPSA_3$	$PPSA_3 = \sum_i asa_i \cdot q_i$, where the summation is limited to atoms with a positive partial charge*: $asa_i \cdot q_i > 0$ (Å^2**e** mol/cm^3)
$PNSA_3$	$PNSA_3 = \sum_i asa_i \cdot q_i$ where the summation is limited to atoms with a negative partial charge: $asa_i \cdot q_i < 0$ (Å^2**e** mol/cm^3)

* Partial charges according to Gasteiger–Marsili were calculated using the rdkit.Chem.rdPartialCharges module.

** Here “e” means the electron charge unit.

Compared to the previous versions of this method, the new tools of MMM presented in this article allowed to build bigger molecules and more complex conformations, which address not only the interaction of adjacent atoms in a polymer chain but also that of atoms came close to each other due to the bending of a macromolecule chain (see Fig. 2.1).

Also, a new improved index calculation algorithm was implemented, which excludes from consideration the outer surface of the oligomer, and thus takes into account interactions of a penetrant with the polymer macromolecule.

The macromolecules in Fig. 2.1 are rather short polymer chains, and their outer shell has considerable surface area. Realistic polymer matrices are much bigger and have no such outer shell. The contribution of atoms that form these outer shells distorts the value of surface-area-based indices from Table 2.1, and their contributions must be excluded from index calculation. Therefore, based on the center-of-mass position of the molecule and the value of the current “probe” radius R, the indices from Table 2.1 were calculated only for atoms of the macromolecule located at no more than $d = 10 - R$ angstrom from its center of mass. Hence, the atoms that belong to the outer shell of the conformation were excluded.

To draw the dependences $ASA(R)$, $ASA^+(R)$, etc., the index value was calculated for each R from 0Å to 0.2Å with a step of 0.05Å, from 0.2Å to 1Å with a step of 0.1Å and from 1Å to 2Å with a step of 0.2Å. Typical dependencies of a number of geometric indices from Table 2.1 from the “probe” radius for one of the polymers, calculated for different d, are presented in Fig. 2.3. One can see from Fig. 2.3 that curves for d = 7Å – R resemble those for d = 10Å – R, so the adopted approach is robust with respect to small deviations of the selection rule for “internal” atoms.

To justify the sufficient chain length and the number of generated conformations for the stability of the results obtained below, 6 conformations of different lengths (200, 300, 400, 600, 800, atoms without hydrogen) were generated for 40 different polymers. For them, indices from the entire set were calculated for a number of R values from the range from 1 Å to 3 Å. In Fig. 2.4 presents, as an example, the obtained dependences of three indices (at different R) on the macromolecule size for 16 random polymers from 40 calculated. It can be seen from the figure that the indices begin to stabilize when the length of the macromolecule exceeds 600 atoms, which was taken as the base length.

The same logic shows that averaging over 10 conformations is enough to obtain stable index values (see Fig. 2.5).

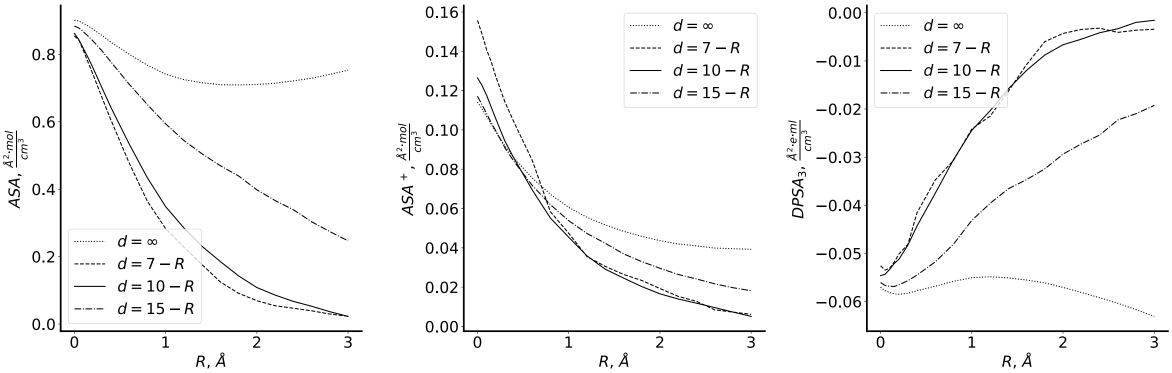

Fig. 2.3. Typical dependencies of a number of geometric indices from Table 2.1 from the "probe" radius for one of the polymers, calculated for various d

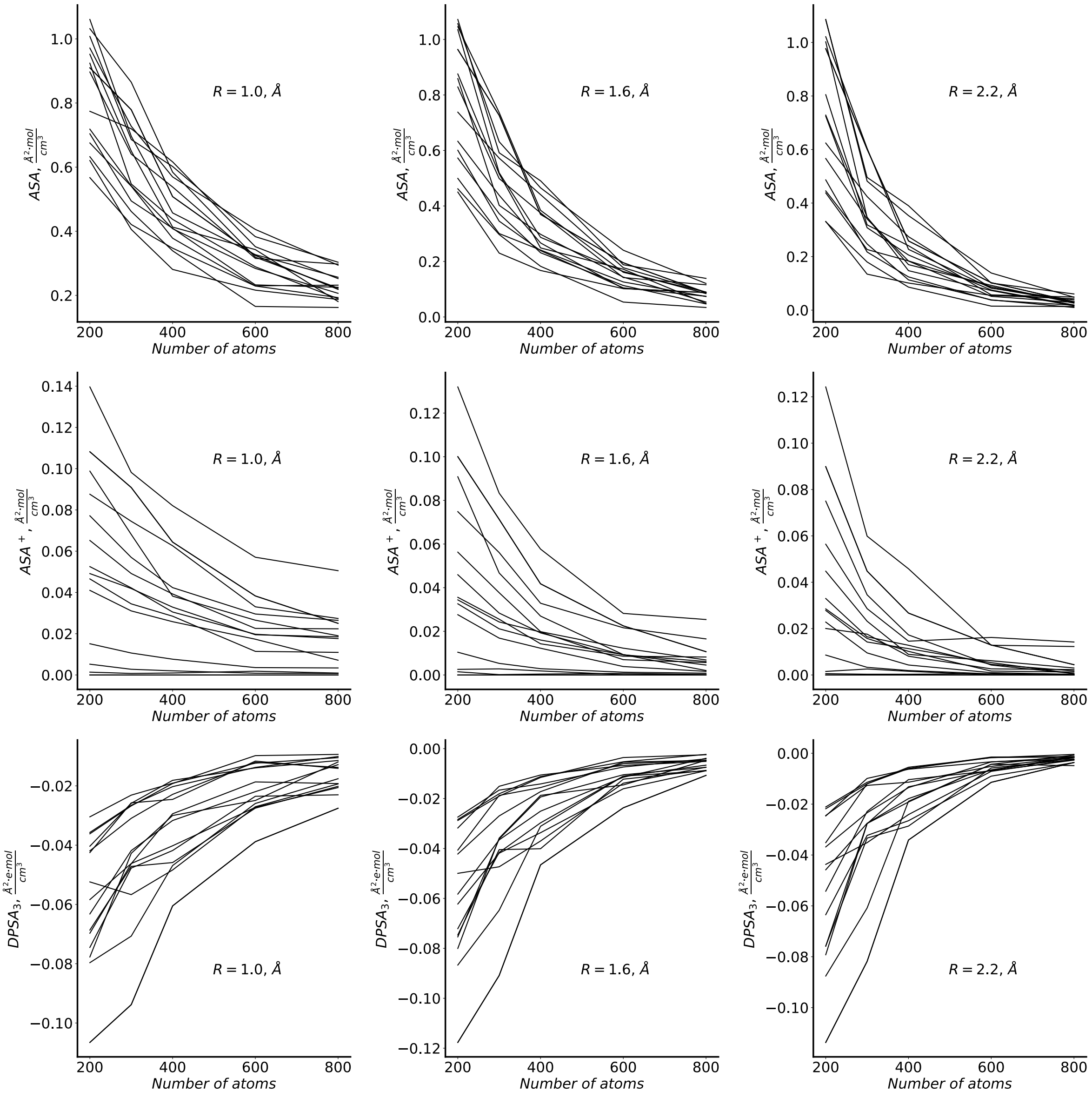

Fig. 2.4. *ASA*, *ASA*$^+$ and *DPSA*$_3$ indices *vs* the macromolecule size for 16 random polymers

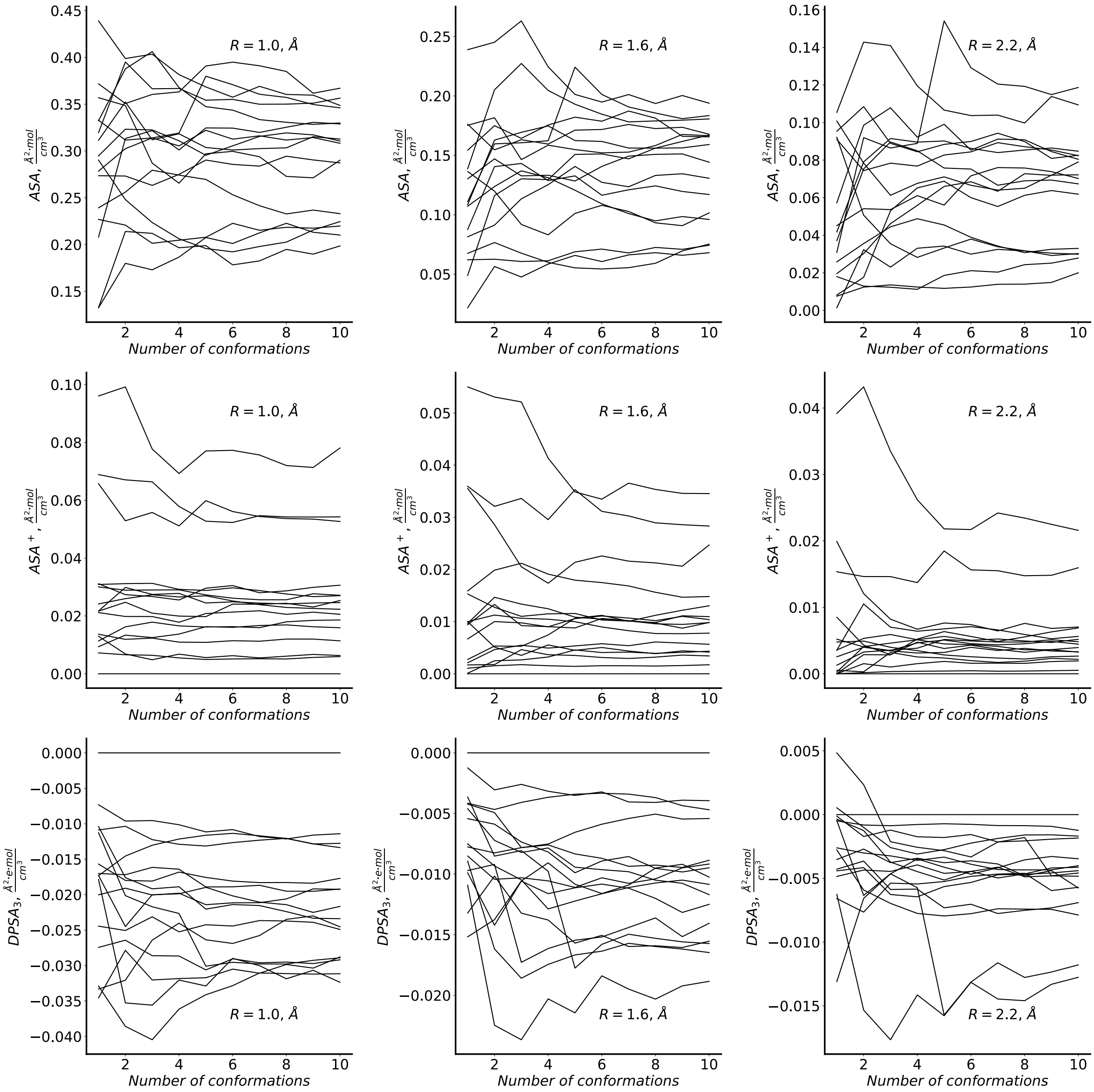

Fig. 2.5. ASA, ASA^+ and $DPSA_3$ indices averaged over conformations *vs* the number of conformations being averaged for 16 random polymers

3. CLASSIFICATION OF CONFORMATIONAL STRUCTURES

In machine learning problems correctly collected and pre-processed data play a key role in obtaining a qualitative result. The database "Gas separation parameters of glassy polymers" (hereinafter referred to as the Database) was created in 1998 at the Laboratory of Membrane Gas Separation at the Topchiev Institute of Petrochemical Synthesis, Russian Academy of Sciences. The database is a unique source of information and a tool for predicting gas separation properties of polymers. Since its inception, the Database is being continuously updated.

The main transport characteristics of polymer gas separation membranes are coefficients of permeability, diffusion, and solubility. The permeability coefficient P can be written as a product of the solubility coefficient S and the diffusion coefficient D. The first (S) describes the driving force of the process of transporting gas molecules, while the second (D) corresponds to the kinetic component of the process.

To solve the clustering problem, 397 unique polymers were selected from the Database, each having several records containing experimental data on the interaction of the gas-polymer pair under different conditions (2629 records in total). While the Database covers more unique polymer structures, for the clustering problem it is necessary to maintain a balance in the data on their chemical classes. Therefore, the structures were selected according to the rule:

- if more than 42 structures of a certain chemical class exist in the database, then records that have experimental data for *P*, *D*, *S* were selected,
- otherwise, all structures of this chemical class were selected.

As justified in Section 2 above, for each polymer in a sample ten conformations were generated of size at least 600 atoms (not counting hydrogen atoms). See Fig. 2.1 for the examples of conformations built.

Then, for each conformation and for each *R* in the range from 0Å to 3Å, the indices from Table 2.1 are calculated using the selection rule $d = 10\text{Å} - R$, which rules off the atoms of the outer hull of the conformation. Typical dependences of geometric indices *ASA*, ASA^+ and $DPSA_3$ on the "probe" radius for one of the polymers are shown in Fig. 3. In contrast to [10], where the optimal range of $[R^-, R^+]$ linearization of the obtained dependencies was searched for, the obtained slope and bias coefficients were used as an explanatory regression variables, in this article the entire curve is used for classification.

For the clustering of polymers, the agglomerative method (a sort of hierarchical classification) was used [1]. The agglomerative clustering method was launched for the number of clusters from $k = 2, ..., 30$, since the optimal number of clusters was unknown. According to the values of the silhouette coefficient, the Calinski–Harabasz index (C-H) and the Davies–Bouldin index (DBI) for the clustering constructed, the value k = 15 has been chosen. The data supplied to the input of the algorithm was previously standardized.

To illustrate the quality of the resulting clustering using the TSNE algorithm [20], the placement of points in 2D space was constructed. It is clearly seen that the TSNE algorithm, being different from agglomerative clustering in its nature, distributes the data points in accordance with the obtained clustering. The embedding also allows to evaluate the distance between clusters (see Fig. 3.1).

Table 3.1. The distribution of polymers of different chemical classes over clusters

Chemical class	**Cluster number**															**Sum by class**
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
polyacrylates	1	3	1	9		1										15
polyesters		2	18	1	1						4		4	4	8	42
polyethers		1	8		12	4	2		2	5			3			37
polyphosphasenes								3								3
polyacetylenes					1	4		3				11				19
polynorbornenes			5	7	2	2	2	9	1	6			3			37
polysulphones					19	1				18					2	40
other N-containing		4	1	5	3						4		9			26
polyamides			11	1	16	3			2				4			37
polystyrenes	2		3	1	2	1		13								22
vinylic polymers	1			3		2	1	5	1				1			14
polycarbonates			14			1							1		3	19
polyimides		3	3	5	6						9		16			42
polyamidoimides		2		2	1		1				11		23			40
other carbo-chained						1				2						3
other hetero-chained								1								1
Sum by cluster	4	15	64	34	63	20	6	34	6	31	28	11	64	4	13	397

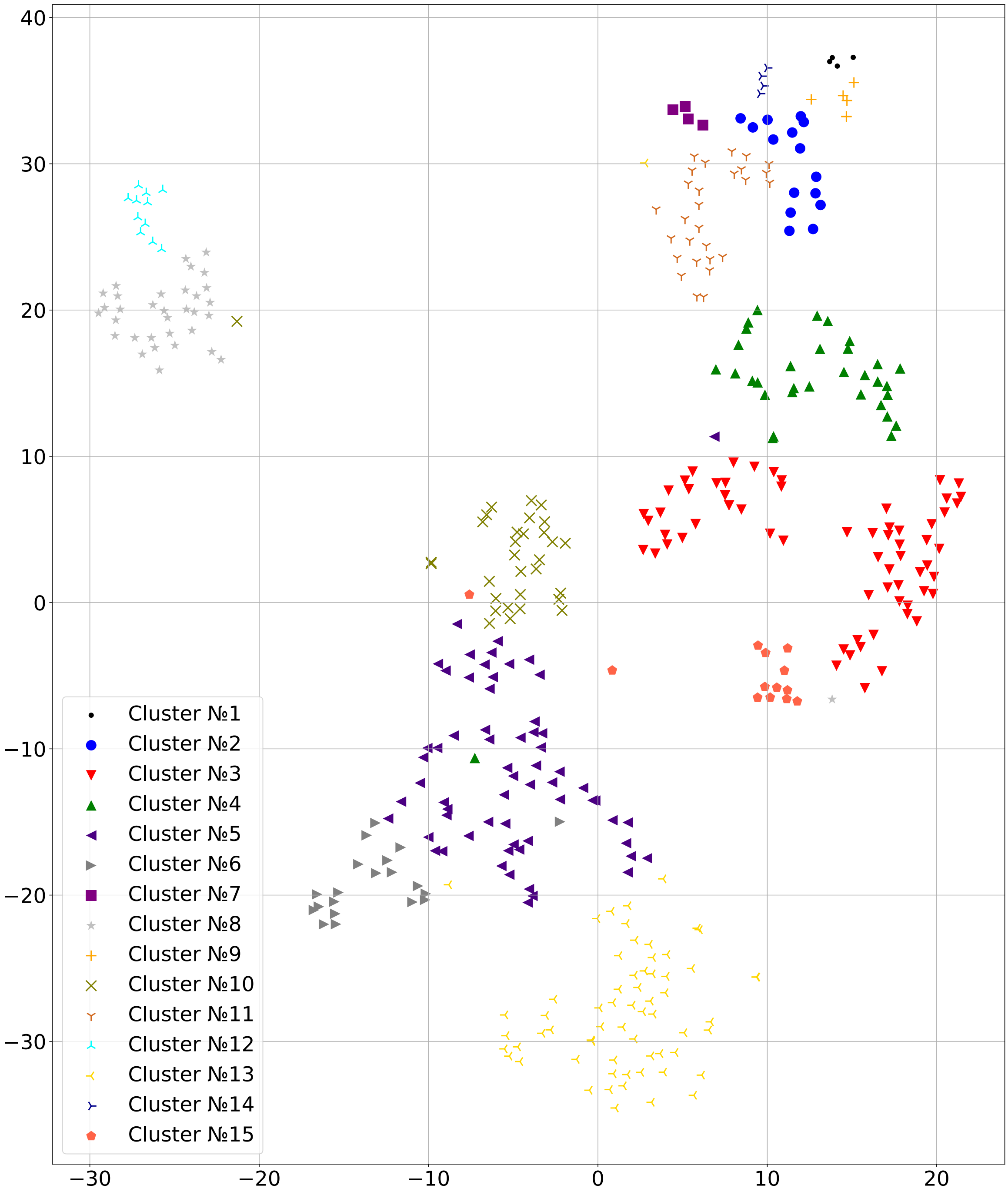

Fig. 3.1. Data embedding in 2D with TSNE algorithm (clusters are shown with different signs and colors)

4. CLUSTERING AND TRANSPORT PROPERTIES

The clusters constructed turn out to be closely related to the transport properties of polymers that are important for membrane technology, first of all, permeability, diffusion, and solubility coefficients of a polymer material with respect to different penetrant gases. The, so-called, Robeson diagram [18] is a conventional representation of transport properties for a collection of polymers. The Robeson diagram is constructed for a pair of penetrant gases (for example, oxygen-nitrogen) in the coordinates $\alpha_P = P_{O2}/P_{N2}$ — separation selectivity of a pair of gases (oxygen-nitrogen) depending on P_{O2} — the permeability coefficient for a more permeable gas (oxygen) and in coordinates $\alpha_D = D_{O2}/D_{N2}$ — diffusion selectivity of a pair of gases (oxygen-nitrogen) depending on D_{O2} — the diffusion coefficient (see Fig. 4.1 and Fig. 4.2, respectively).

Robeson diagrams are often used by membrane gas separation specialists for comparative analysis of different groups of polymers. Hence, it is quite natural to represent the clusters of polymers constructed in the previous section as the points on a Robeson diagram to analyze the relation between the obtained classification of polymers and their transport properties.

The experimental data on the coefficient of permeability and diffusion were taken from the Database. Since the experimental data were obtained at various temperatures, they were brought to a single standard value of 308K using the algorithm from [2].

Then, the diagrams were constructed using the experimental data on the permeability (see Fig. 4.1) and diffusion (see Fig. 4.2) coefficients available from the Database. For the convenience of the analysis, along with the points of an individual cluster the center of this cluster and the center of the entire data sample (calculated without the explicit outliers) are added to the Robeson diagram.

The clusters turn out to be closely related to the transport properties of polymers. The location of the center of mass of the cluster relative to the center of mass of the entire data sample, the shape of the cluster and its composition demonstrate these relationships. The first thing worth noting is that the points of most clusters are rather crowded on the Robeson diagrams. Consequently, materials with similar transport properties are collected in one cluster. It is important that clustering was built only on the basis of the shape and geometry of conformations of polymer molecules without using any hint like the chemical class of a polymer.

Now consider the clusters separately. The most revealing relationships can be found on the Robeson diagram for the diffusion selectivity in Figure 4.2. So, the centers of mass of large clusters 3, 5, 13 are shifted to the upper left quadrant of the general diagram. This indicates a high selectivity for D with a low diffusivity. Moreover, in combination with the information from Fig. 4.1, polymers from these clusters are a good combination of permeability selectivity and permeability itself. Also, cluster 11 can be attributed to them if there were more experimental data on the diffusion coefficient. For the most part, polymers from clusters 6, 8, 12 turned out to be low-selective in terms of permeability and diffusion. However, they were highly permeable and highly diffuse.

From Table 3.1 it can be noted that, in general, clustering does not depend on the chemical class of polymers. So, the above clusters 3, 5, 6, 8, 13 consist of polymers of various chemical classes. However, there are exceptions. For example, all highly diffuse and highly permeable polymers with a low selectivity from 12 cluster are polyacetylenes. All polyphosphasenes entered cluster 8, most polycarbonates entered cluster 3, and polysulphones were divided into two clusters 5 and 10. Polyamidoimides and polyimides, being often similar in structure and properties, were mainly divided into clusters 11 and 13.

Let us turn to the structure of polymers themselves. Cluster 6 consists only of polymers with bulky fluorine substituents, which in this case belong to different chemical classes. In combination with the information that the polymers of this cluster are highly permeable and highly diffuse and, at the same time, low selective, one can judge the consequences of adding bulky fluorine substituents. Moreover, cluster 7, which contains mainly polymers with a pentafluorophenyl group, is rather the opposite of low selectivity and low diffusion. This suggests that not always the presence of a large number of fluorine molecules leads to properties similar to polymers of 6. It is also interesting to compare cluster 6 and cluster 10. Cluster 10 also has bulky fluorine substituents; however, monomer units are often several times longer and most polymers belong to polyamidoimides and polyimides, which are never met in cluster 6.

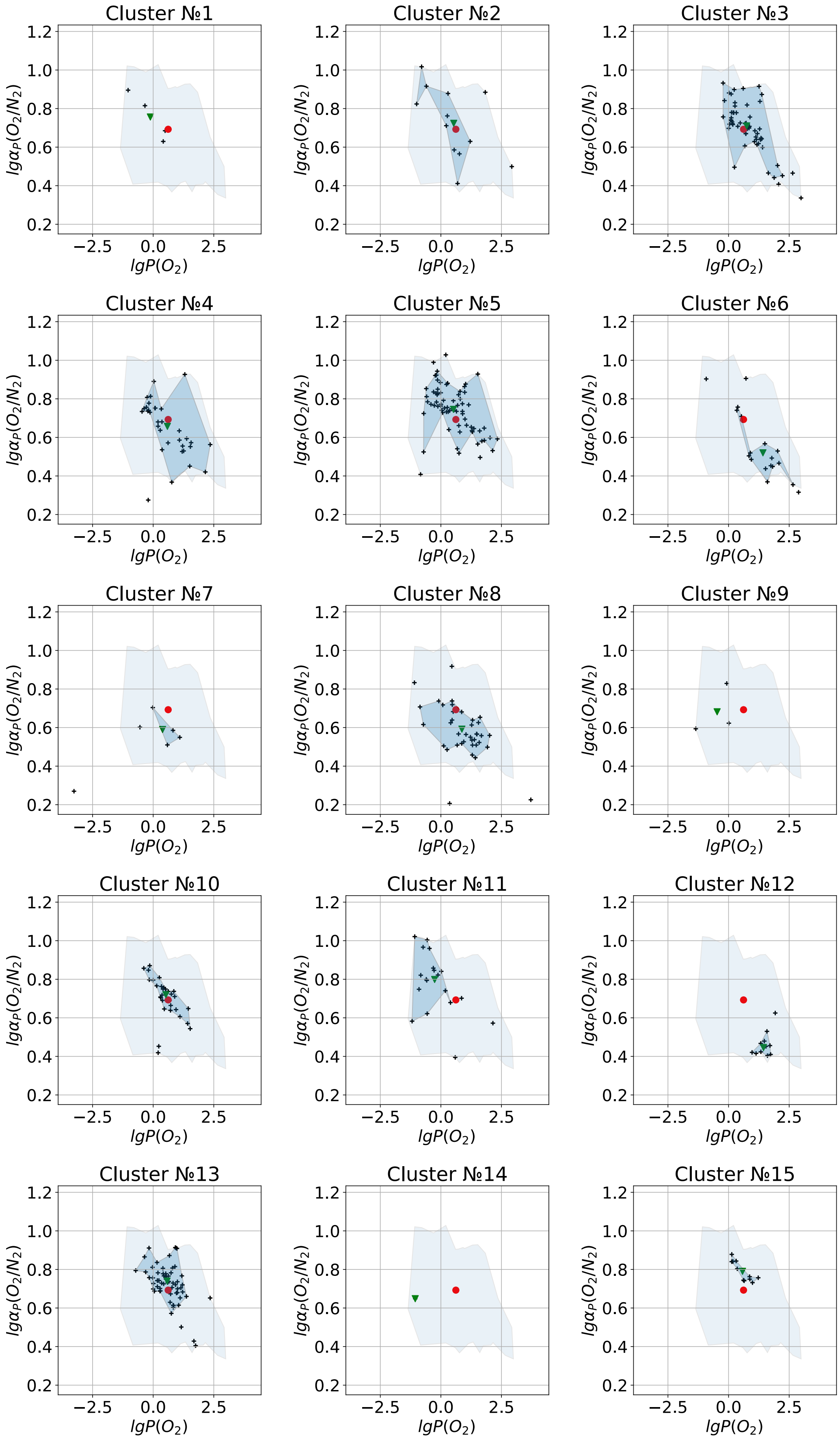

Fig. 4.1. Polymer clusters on the Robeson chart for permeability. The center of the whole data sample is depicted with the red circle, while the center of the cluster is depicted with the green triangle

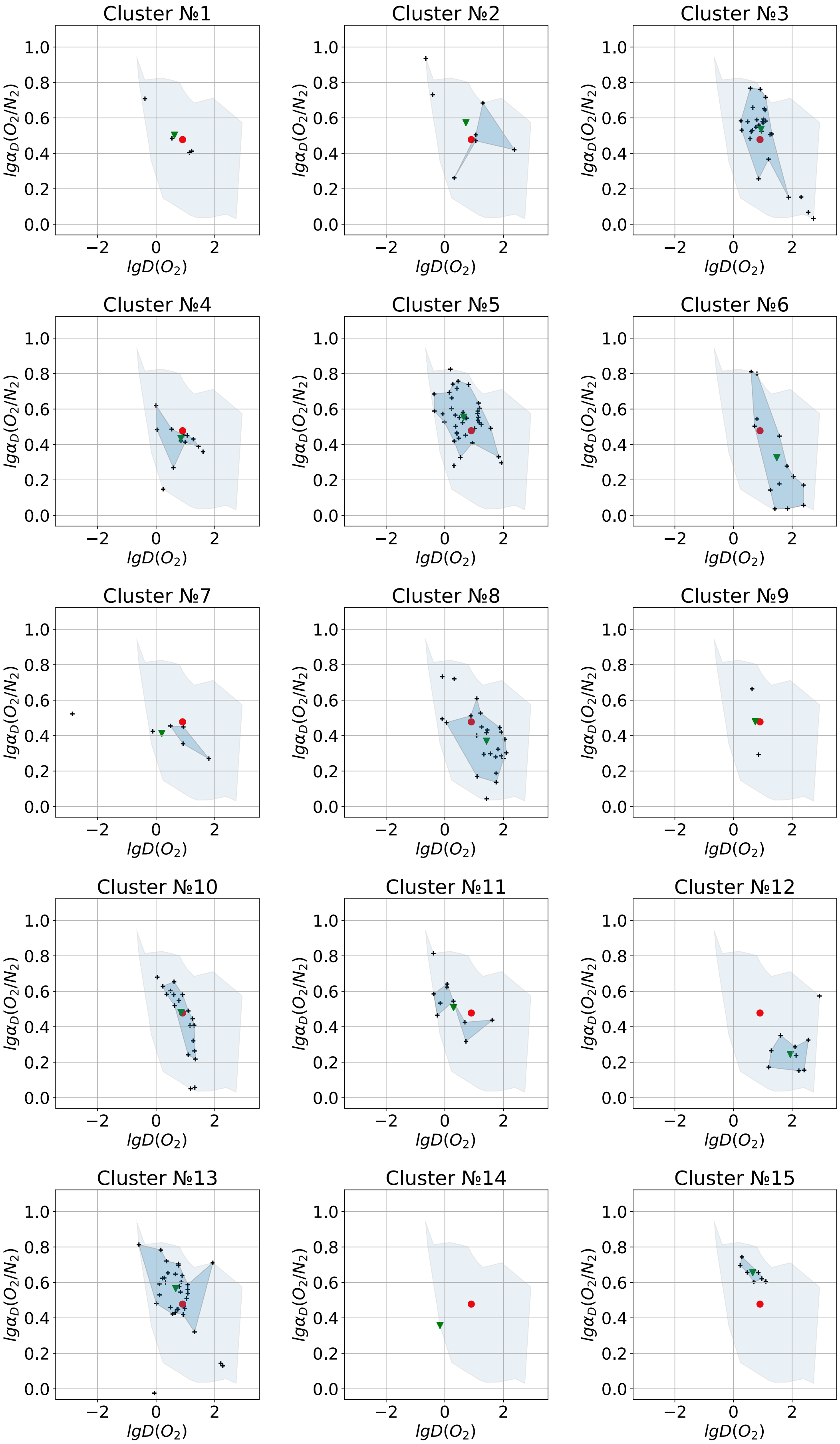

Fig. 4.2. Polymer clusters on the Robeson chart for diffusion. The center of the whole data sample is depicted with the red circle, while the center of the cluster is depicted with the green triangle

The above analysis reveals the deep relation between the transport properties of polymers and the shape and geometry of conformations of their macromolecules. The resulting clustering allows one to identify the signs and characteristics of polymer molecules with various extreme properties, which will undoubtedly be useful in the search and synthesis of new promising polymers.

5. CONCLUSION

Based on a large sample consisting of almost 400 amorphous polymers from 16 different chemical classes used in membrane gas separation, clustering was constructed only on the basis of data on the shape and geometry of the conformations of polymer molecules. Obtained 15 clusters are closely related to transport parameters important for membrane gas separation, such as permeability coefficient P and diffusion coefficient D.

The method proposed consists of several steps. At the beginning, 10 realistic conformations of polymer macromolecules are constructed. For these purposes, a program code is written in the RDKit environment for Python. It satisfies all the requirements listed in chapter 2: the realism and reproducibility of the resulting structures, fast automated calculation, the possibility of parallelization, and free distribution of software. Then, the dependencies were calculated of 8 accessible-surface-area-based indices (see Table 2.1) on the radius of a spherical probe that represents the variety of penetrant gases. An improved algorithm for calculating indices eliminates the influence of the outer shell of the macromolecule, which allows focusing on the processes occurring inside the polymer membrane. The obtained 8 curves are averaged between ten polymer conformations to obtain more robust figures. Then, these dependences characterizing the polymer are used as predictors for the agglomerative clustering method.

The plans for further research include constructing separate regressions for the resulting clusters in order to more accurately predict transport properties (P, D, and S coefficients). The methods developed allow us to approach the solution of the problem of obtaining substances with predetermined properties. For example, the prediction of transport parameters for a collection of hypothetic polymers, followed by the manual selection of several most promising polymers, could guide the synthesis experiments of novel polymers.

ACKNOWLEDGEMENTS

The reported study was funded by RFBR according to research projects 18-37-00265 and 19-37-90004.

REFERENCES

1. Agglomerative Clustering. *Recursively merges the pair of clusters that minimally increases a given linkage distance* [Online]. Available https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html
2. Alent'ev A.Ju. (2003) Prognozirovanie transportnyh svojstv stekloobraznyh polimerov: Rol' himicheskoj struktury i svobodnogo ob'ema. *Doctor of science Thesis*, Moscow [in Russian].
3. Anderson J. A., Lorenz C. D., Travesset A. (2008) General purpose molecular dynamics simulations fully implemented on graphics processing units, *Journal of Computational Physics*, 227(10), 5342-5359.
4. ChemAxon. *Software solutions and services for chemistry & biology* [Online]. Available https://www.chemaxon.com
5. ChemOffice Professional. *Chemical Communications Software* [Online]. Available http://www.perkinelmer.com/Product/chemoffice-professional-chemofficepro
6. Conformer Plugin. *Walk-through manual on how to use the Conformer Plugin* [Online]. Available https://docs.chemaxon.com/display/docs/Conformer+Plugin

7. Eastman P., et al. (2017) OpenMM 7: Rapid development of high performance algorithms for molecular dynamics, *PLOS Comp. Biol*, 13(7), e1005659.
8. Fried, J. (2006) *Materials Science of Membranes for Gas and Vapor Separation*, Ed. by Yu. Yampolskii, I. Pinnau, B.D. Freeman, Wiley, Chichester, p. 95.
9. Glaser J., et.al. (2015) Strong scaling of general-purpose molecular dynamics simulations on GPUs, *Computer Physics Communications*, 192, 97-107, https://doi.org/10.1016/j.cpc.2015.02.028
10. Goubko M., Miloserdov O., Yampolskii Yu., Alentiev A., Ryzhikh V. (2019) Prediction of Solubility Parameters of Light Gases in Glassy Polymers on the Basis of Simulation of a Short Segment of a Polymer Chain, *Polymer Science*, 61(5), 718–732.
11. Goubko M., Miloserdov O., Yampolskii Yu., Alentiev A., Ryzhikh V. (2016) A novel model to predict infinite dilution solubility coefficients in glassy polymers, *Journal of Polymer Science Part B: Polymer Physics*, 55(3), 228-244.
12. LAMMPS. *Molecular Dynamics Simulator* [Online]. Available https://lammps.sandia.gov/
13. Norman, G.E.; Filinov, V.S. (1969) *High Temp*, 7, 216.
14. Rappe A. K. et.al. (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, *Journal of the American Chemical Society*, 114 (25), 10024-10035.
15. rdDistGeom module. *Module containing functions to compute atomic coordinates in 3D using distance geometry* [Online]. Available https://www.rdkit.org/docs/source/rdkit.Chem.rdDistGeom.html
16. RDKit. *Open-Source Cheminformatics Software* [Online]. Available http://www.rdkit.org
17. Robeson L.M. et. Al. Membr. Sci., 1997, 132, 33.
18. Robeson, L.M. (1991) Correlation of separation fac-tor versus permeability for polymeric membranes, *Journal of Membrane Science*, 62 (165), 165–185.
19. Theodorou, D. (2006) *Materials Science of Membranes for Gas and Vapor Separation*, Ed. by Yu. Yampol-skii, I. Pinnau, B.D. Freeman, Wiley, Chichester, p. 49.
20. TSNE. *T-distributed Stochastic Neighbor Embedding* [Online]. Available https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html
21. Yampolskii Yu., Shishatskii S., Alentiev A., Loza K. (1998), *J. Membr. Sci*. 148, 59.