Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Tuning the Molecular Electronic Structure of Polysaccharides: A Comparative DFT Study of Chemically Modified Chitosan and Cellulose Derivatives

Nilufar Burkhanova
Abdulla Kadiry st
Ziyoda Abrarova
Abdulla Kadiry st
Akmal Abilkasimov
Kimyo international university in Tashkent
Ilnar Nurgaliev
ilnarvodnik@gmail.com

Abdulla Kadiry st

Research Article

Keywords: chitosan, cellulose, frontier molecular orbitals, HOMO–LUMO gap, conceptual DFT, unsupervised machine learning, polysaccharide functionalization

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11440857/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Tuning the Molecular Electronic Structure of Polysaccharides: A Comparative DFT Study of Chemically Modified Chitosan and Cellulose Derivatives

Nilufar Burkhanova[1], Ziyoda Abrarova[1], Akmal Abilkasimov[2] and Ilnar Nurgaliev[1]

[1]Theoretical Department, Institute of polymer chemistry and physics, Abdulla Kadiry st., Tashkent, 100128, State, Uzbekistan.

[2]Department of Chemistry and Biology, Kimyo international university in Tashkent, Shota Rustaveli st., Tashkent, 100121, State, Uzbekistan.

Corresponding author E-mail: ilnarvodnik@gmail.com;

Contributing authors: nburhanova1707@gmail.com;

Abstract

The rational molecular design of functionalized polysaccharides for advanced electronic and sensing applications necessitates a comprehensive mapping of their electronic chemical space. In this study, we report a high-throughput computational screening of 115 chemically modified chitosan and cellulose derivatives. Frontier molecular orbital (FMO) energies, HOMO–LUMO energy gaps (ΔE_{gap}), and a comprehensive suite of global conceptual density functional theory (DFT) descriptors were computed for isolated oligomeric models at the B3LYP/6-31G(d,p) level with SMD implicit solvation. Our quantum chemical analysis reveals that while native polysaccharide backbones exhibit wide ΔE_{gap} values (5.8–6.1 eV), characteristic of electronically rigid systems, targeted functionalization with π-conjugated aromatics, highly polarizable chalcogens (S, Se), and transition-metal coordination centers systematically narrows the gap to 2.6–4.5 eV, significantly enhancing intramolecular charge-transfer capabilities. Furthermore, chitosan derivatives demonstrate superior electronic tunability compared to cellulose due to the reactive C-2 amino group. To navigate the resulting high-dimensional descriptor space, we implemented an unsupervised machine-learning pipeline integrating Principal Component Analysis (PCA), K-Means clustering, and Isolation Forest anomaly detection. This chemoinformatics approach successfully delineated distinct electronic clusters and identified transition-metal complexes and specific dye conjugates as statistical outliers with exceptionally narrow energy gaps. Ultimately, the generated molecular electronic-property map provides a robust, data-driven framework for the rational computer-aided design of next-generation functional polysaccharide materials.

Keywords: chitosan, cellulose, frontier molecular orbitals, HOMO–LUMO gap, conceptual DFT,

unsu-pervised machine learning, polysaccharide functionalization.

1 Introduction

The rational design of functional macromolecular materials with tailored electronic properties requires a comprehensive mapping of their electronic chemical space. Naturally occurring polysaccharides, particularly chitosan and cellulose, represent a highly attractive and structurally diverse chemical space for molecular engineering. They possess reactive hydroxyl and amino functional groups that are amenable to extensive chemical modification, allowing for the systematic tuning of their physicochemical characteristics [1–9]. While historical research on these biopolymers has been predominantly confined to macroscopic physicochemical or biological applications [10–15], their emerging utility in molecular electronics, electrochemical sensing, and energy-storage devices necessitates a fundamental, quantum-chemical understanding of their electronic structure [1, 2, 5, 6]. Realizing this potential requires moving beyond empirical trial-and-error synthesis toward the rational, computer-aided modulation of molecular electronic properties. Strategies involving the targeted introduction of aromatic π-conjugated fragments, highly polarizable heteroatoms (such as sulfur and selenium), and transition-metal coordination sites have been shown to profoundly influence molecular polarization, orbital delocalization, and intramolecular charge-transfer dynamics [3, 7–9].

Despite significant progress in polysaccharide research, a systematic, high-throughput computational understanding of the electronic structure of chemically modified derivatives remains limited. While recent experimental studies continue to emphasize bulk structural characterization and biological applications [16], computational investigations have increasingly employed atomistic simulations and in silico design to elucidate crystal structures, dissolution mechanisms, and targeted nanoformulations [17, 18]. Furthermore, density functional theory (DFT) and molecular dynamics (MD) have been successfully applied to gain electrical insights into biopolymer electrolytes [19] and to model specific biopolymer–additive interactions at the molecular level [20]. However, these computational efforts are typically restricted to ad hoc analyses of isolated systems or specific composite formulations. Crucially, there is an absence of a unified, high-throughput computational database that systematically evaluates frontier molecular orbital (FMO) characteristics and global reactivity descriptors across a broad, chemically diverse spectrum of modified polysaccharides. Crucially, there is an absence of a unified, high-throughput computational database that systematically evaluates the frontier molecular orbital (FMO) characteristics and global reactivity descriptors across a broad, chemically diverse spectrum of modified polysaccharides.

Frontier molecular orbital (FMO) theory and conceptual DFT provide a rigorous theoretical framework for the in silico evaluation of molecular electronic behavior. The energies of the highest occupied and lowest

unoccupied molecular orbitals (E_{HOMO} and E_{LUMO}), along with the HOMO–LUMO energy gap (ΔE_{gap}), serve as fundamental descriptors of electron-donor ability, electron affinity, kinetic stability, and intramolecular charge-transfer capability [21]. Furthermore, these parameters form the basis for global conceptual DFT indices—such as chemical hardness (η), global softness (S), chemical potential (μ), and electrophilicity index (ω)—which are increasingly utilized for the rational computational screening and quantitative structure-property relationship (QSPR) modeling of molecular reactivity [22–24].

The integration of chemoinformatics and machine learning (ML) has transformed computational molecular design by enabling the efficient navigation and dimensionality reduction of vast chemical spaces [25–29]. While supervised ML models are often employed for direct property prediction, unsupervised learning techniques offer distinct advantages for exploratory data analysis, the identification of hidden structural patterns, and the detection of statistical anomalies within high-dimensional descriptor datasets. Nevertheless, the application of unsupervised ML workflows to systematically map the electronic chemical space of chemically modified chitosan and cellulose remains largely unexplored.

The present study addresses these critical gaps by establishing a comprehensive computational screening dataset comprising 115 chemically modified derivatives of chitosan and cellulose. We systematically compute FMO energies, HOMO–LUMO gaps, and global conceptual DFT descriptors to elucidate the influence of diverse substituent classes—including electron-donating/withdrawing groups, conjugated aromatic systems, polarizable heteroatoms, and transition-metal complexes—on the intramolecular electronic structure of these biopolymers. Furthermore, we implement an unsupervised machine-learning pipeline, incorporating Principal Component Analysis (PCA), K-Means clustering, and Isolation Forest anomaly detection, to map the multidimensional descriptor space and identify statistically distinct electronic clusters.

The scientific novelty of this work resides in: (i) the construction of one of the most extensive molecular electronic descriptor datasets for modified polysaccharides, enabling robust chemoinformatic analysis; (ii) the systematic comparative evaluation of diverse functional-group effects on molecular polarizability and charge-transfer capability via conceptual DFT; and (iii) the application of an unsupervised ML framework to navigate the polysaccharide chemical space. Ultimately, this data-driven approach identifies high-priority molecular candidates and provides a theoretical foundation for the rational, computer-aided design of next-generation functional biopolymer materials.

2. Computational Details and Methodology

2.1. Dataset Construction and Molecular Modeling

A comprehensive computational dataset was assembled to investigate the influence of chemical functionalization on the electronic properties of polysaccharide-based materials. The dataset comprises 115 chemically modified derivatives of chitosan and cellulose, collected from published literature describing established synthetic routes. The selected compounds represent the major classes of polysaccharide functionalization currently employed in biomedical engineering, environmental technologies, catalysis, sensing, and sustainable electronic materials [30–36]. Specifically, the analyzed structures encompass amino-functionalized, hydroxyalkylated, alkylated, carboxylated, sulfated, phosphorylated, aromatic, azido-containing, sulfur-containing, selenium-containing, halogenated, and transition-metal-coordinated derivatives.

To ensure computational feasibility for the high-throughput screening of 115 distinct structures, each chemically modified polysaccharide was modeled as an isolated representative oligomeric unit (typically a monomeric or dimeric repeat unit, depending on the specific modification site and steric requirements). The convergence behavior of the HOMO–LUMO energy gap with respect to the degree of polymerization was systematically evaluated, confirming that the electronic properties stabilize already at the dimeric or trimeric level (see Figure S1 in Supporting Information). It is crucial to emphasize that the calculated frontier orbital properties describe the intrinsic intramolecular electronic characteristics of the isolated chains in a solvated environment. акгзэўThese molecular descriptors serve as proxy parameters for evaluating the material's molecular polarizability and intramolecular charge-transfer capability, rather than direct measures of solid-state electronic conductivity or bulk bandgap. For internal consistency, each derivative was treated as an independent molecular system characterized by a common set of electronic descriptors. The complete dataset of the 115 investigated derivatives, including the calculated frontier molecular orbital energies and the full set of global conceptual DFT descriptors, is provided in Table S1 in the Supporting Information

2.2. Quantum Chemical Calculations

All molecular geometries were optimized using density functional theory (DFT) as implemented in Gaussian 16 Rev. C.01 [37]. The hybrid B3LYP exchange–correlation functional [38] combined with the 6-31G(d,p) basis set [39] was selected. While the B3LYP functional is known to systematically underestimate HOMO–LUMO gaps in extended ππ-conjugated systems and transition-metal complexes due to self-interaction errors, it was chosen for this study to maintain a balanced description of hydrogen bonding and covalent interactions across a highly diverse dataset, while keeping computational costs tractable for 115 systems. The relative trends across the dataset remain physically meaningful for comparative high-throughput screening. For structures containing transition metals (Co, Ti, V complexes), the def2-SVP basis set [40] together with the corresponding effective core potentials (ECPs) was adopted. Single-point energy calculations on the

optimized geometries were subsequently refined using the larger def2-TZVP basis set [49] to improve the accuracy of the frontier orbital energies.

Geometry optimization was performed without symmetry constraints using tight self-consistent field (SCF) convergence criteria. The SCF energy convergence criterion was set to 10−810−8 Hartree. To verify that each optimized structure corresponds to a true minimum on the potential-energy surface, harmonic vibrational frequency calculations were performed at the same level of theory; the absence of imaginary frequencies confirmed successful optimization in all 115 cases.

Because most chitosan and cellulose derivatives are synthesized and applied under aqueous conditions, implicit solvent effects were included using the Solvation Model based on Density (SMD) [41] with water as the solvent (ε=78.4). The SMD model improves the description of molecular polarization and hydrogen-bond stabilization, yielding frontier orbital energies that are more representative of experimental conditions than gas-phase calculations.

2.3. Conceptual DFT Descriptors

The electronic properties of all derivatives were evaluated using frontier molecular orbital (FMO) descriptors. The principal parameters were the energy of the highest occupied molecular orbital (E_{HOMO}), the energy of the lowest unoccupied molecular orbital (E_{LUMO}), and the HOMO–LUMO energy gap (ΔE_{gap}):

$$\Delta E_{gap} = E_{LUMO} - E_{HOMO} \quad (1)$$

Note: To avoid confusion with the solid-state bandgap of bulk materials, the term ΔE_{gap} is used exclusively to denote the molecular excitation energy gap.

Based on Koopmans' approximation, global conceptual DFT descriptors were calculated for every derivative. The ionization potential (IP) and electron affinity (EA) were estimated as:

$$IP = -E_{HOMO} \quad (2)$$

$$EA = -E_{LUMO} \quad (3)$$

It should be noted that Koopmans' theorem neglects orbital relaxation and electron correlation effects. Thus, while IP and EA values are excellent for relative ranking within the dataset, their absolute values for systems with strong electronic correlation, such as the transition-metal complexes, should be interpreted with caution.

The chemical hardness (η) was calculated as:

$$\eta = (IP - EA)/2 \quad (4)$$

The chemical potential (μ) was determined using:

$$\mu = -(IP + EA)/2 \quad (5)$$

The global softness (S) was calculated as:

$$S = \frac{1}{2\eta} \quad (6)$$

Finally, the electrophilicity index (ω) was obtained from:

$$\omega = \frac{\mu^2}{2\eta} \quad (7)$$

Spatial distributions of HOMO and LUMO orbitals were visualized using isosurfaces generated at an isovalue of 0.02 a.u. Molecular electrostatic potential (MEP) maps were calculated on an electron density isosurface (0.001 e $bohr^{-3}$). Atomic charge distributions were analyzed using Natural Bond Orbital (NBO) analysis to evaluate donor–acceptor interaction energies.

2.4. Chemoinformatics and Unsupervised Machine Learning Workflow

To investigate the electronic descriptor space of the chitosan and cellulose derivatives without predefined class labels, an unsupervised machine learning workflow was implemented using Python 3.11 and the scikit-learn library [42]. The workflow consisted of data preprocessing, dimensionality reduction via Principal Component Analysis (PCA), unsupervised clustering using the K-Means algorithm, and anomaly detection employing the Isolation Forest method.

Prior to multivariate analysis, all conceptual DFT descriptors (E_{HOMO}, E_{LUMO}, ΔE_{gap}, IP, EA, μ, η, S, and ω) were standardized to zero mean and unit variance using the StandardScaler algorithm. This ensures that each variable contributes equally to the subsequent statistical analyses, eliminating biases arising from differences in descriptor magnitude.

Principal Component Analysis (PCA) was employed to reduce the dimensionality of the standardized descriptor matrix. PCA transforms the original correlated variables into orthogonal principal components representing linear combinations of the original descriptors while maximizing the explained variance. The first two principal components were retained for visualization, as they accounted for the vast majority of the total variance. Loading vectors were analyzed to determine the contribution of each conceptual DFT descriptor to the principal components.

K-Means clustering was applied to identify groups of derivatives with similar electronic characteristics. Cluster centroids were initialized using the K-means++ procedure. The algorithm was executed with 100 independent initializations (ninit=100), a maximum of 500 iterations (max_iter=500), and random_state=42 to ensure computational reproducibility. The optimal number of clusters was selected by simultaneously considering the Silhouette coefficient, Davies–Bouldin index, and Calinski–Harabasz criterion.

The overall computational strategy, integrating quantum chemical calculations with unsupervised machine learning workflows, is summarized in Figure 1. The separation along PC1 primarily reflects

variations in global reactivity, with highly reactive derivatives located at negative PC1 values and kinetically stable compounds occupying the positive side of the axis. In contrast, PC2 mainly differentiates molecules according to their ionization potential and HOMO energy, providing an additional level of discrimination between electron-rich and electron-deficient derivatives. The clear spatial separation of the clusters indicates that functionalization of the polysaccharide backbone systematically modifies the electronic structure and creates chemically meaningful classes of materials.

Potential electronic outliers were identified using the Isolation Forest algorithm, an ensemble tree-based method designed for unsupervised anomaly detection. The model was constructed using 300 isolation trees (nestimators=300) with a contamination factor of 0.05 (corresponding to the expected proportion of anomalous observations) and random_state=42. Molecules exhibiting significantly shorter average isolation path lengths than the remaining derivatives were classified as electronic anomalies.

The quality of the clustering solution was quantitatively evaluated using three complementary internal validation metrics: the Silhouette coefficient (assessing cluster cohesion and separation), the Davies–Bouldin index (evaluating average similarity between neighboring clusters), and the Calinski–Harabasz index (quantifying the ratio between inter-cluster and intra-cluster variance). The results of these validation metrics are summarized in Table 1.

Table 1 Internal validation metrics of the K-Means clustering solution.

Metric	Value	Interpretation
Silhouette score	0.478	Moderate-to-good cluster separation and cohesion
Davies–Bouldin Index	0.717	Low inter-cluster similarity (lower values indicate better clustering)
Calinski–Harabasz Index	140.99	Well-defined cluster structure (higher values indicate better separation)

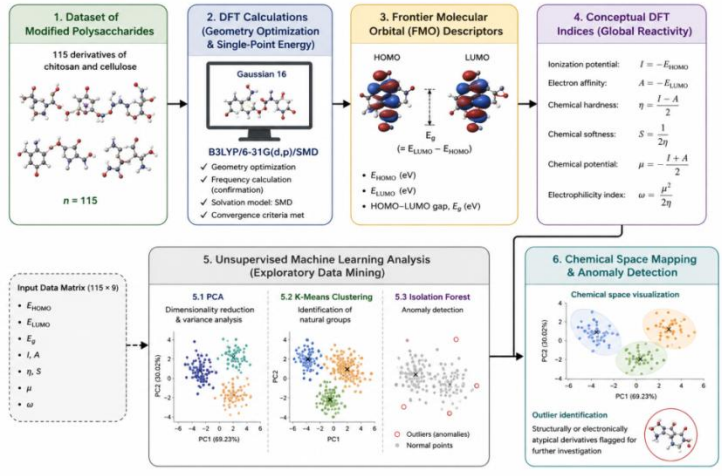


Fig. 1 Schematic workflow diagram illustrating the computational strategy: from molecular modeling of 115 chitosan and cellulose derivatives through DFT calculations, conceptual descriptor analysis, to unsupervised machine learning (PCA, K-Means clustering, and Isolation Forest anomaly detection).

3 Results and Discussion

3.1. Electronic characteristics of chitosan and cellulose derivatives

3.1.1. General trends in frontier molecular orbitals

The calculated frontier molecular orbital (FMO) energies demonstrate that the electronic properties of the investigated polysaccharide models are predominantly governed by the nature of the introduced functional groups rather than by the carbohydrate backbone itself. For the native chitosan and cellulose models, the calculations yield wide HOMO–LUMO energy gaps (ΔE_{gap}) in the range of 5.8–6.1 eV. These large gaps are characteristic of electronically rigid molecular systems with high kinetic stability and low intramolecular charge-transfer capability.

Upon chemical functionalization, the HOMO energies shift from –6.6 to –4.9 eV, and the LUMO energies vary between –2.5 and –0.2 eV, resulting in a broad distribution of ΔE_{gap} values (2.6–5.6 eV). This wide variation confirms that rational molecular engineering provides an effective strategy for tuning the intramolecular electronic structure, molecular polarizability, and charge-transfer characteristics of polysaccharides over a broad energetic range.

3.1.2. Influence of aliphatic, hydroxyl, and ionic substituents

Derivatives functionalized with saturated aliphatic chains (e.g., heptyl-, octyl-, palmitoyl-, and stearoyl-chitosan), hydroxyl groups (e.g., hydroxyethyl- and hydroxypropyl-derivatives), or alkyl-substituted amino groups (e.g., diethylaminoethyl-chitosan) exhibit relatively minor perturbations in their frontier electronic structure. These compounds maintain wide HOMO–LUMO gaps ($\Delta E_{gap} \approx 5.1–5.6$ eV).

The underlying mechanism is primarily inductive: electron-donating alkyl and hydroxyl groups exert a weak positive inductive (+I) effect, which slightly destabilizes the HOMO (raising its energy) but does not introduce new delocalized electronic states near the LUMO. Consequently, the frontier orbitals remain localized on the glucosamine or glucopyranose framework. Similarly, strongly electron-withdrawing ionic groups (e.g., carboxylate, sulfonate, and sulfate) stabilize both HOMO and LUMO levels through negative inductive effects, yielding deep HOMO energies (down to –6.4 eV) and enhanced oxidative stability. However, because both frontier orbitals shift synchronously, the ΔE_{gap} remains relatively large (4.8–5.2 eV). These findings indicate that purely aliphatic, hydroxyl, or ionic modifications primarily modulate the chemical robustness, amphiphilicity, and ion-coordination capability of the polymers, rather than fundamentally enhancing their intramolecular charge-transfer capability.

3.1.3. Impact of π-conjugation and heteroatom polarizability

A substantially different electronic response is observed upon the introduction of conjugated aromatic systems and highly polarizable heteroatoms. The incorporation of phenolic, azo, thiosemicarbazide, and Schiff base (imine) fragments significantly narrows the ΔE_{gap} (to the 3.7–4.8 eV range) through π-electron delocalization. The formation of extended conjugated networks (e.g., C=N or C=C–C=O) facilitates intramolecular charge transfer by stabilizing the LUMO through π* antibonding orbital formation.

The most pronounced reduction in ΔE_{gap} among purely organic derivatives is achieved by sulfur- and selenium-containing functionalities. Dithiocarbamate-, phenylthiocarbamoyl-, and seleno-chitosan derivatives exhibit ΔE_{gap} values as low as 3.2–3.5 eV. This remarkable behavior originates from the unique electronic structure of chalcogens: sulfur and selenium possess highly polarizable 3p and 4p valence orbitals, respectively, which overlap efficiently with adjacent π orbitals. This overlap promotes extensive electron delocalization and introduces low-energy antibonding orbitals, thereby substantially stabilizing the LUMO while simultaneously raising the HOMO energy. The resulting highly polarized electronic structures suggest that these heteroatom-functionalized polysaccharides possess enhanced molecular softness and are highly promising for applications requiring efficient intramolecular charge redistribution, such as electrochemical sensing and redox-active biointerfaces.

The systematic influence of functional group classes on the HOMO–LUMO energy gap is summarized in Fig. 2, which clearly illustrates the progressive narrowing of ΔE_{gap} upon introduction of conjugated systems, polarizable chalcogens, and transition metals.

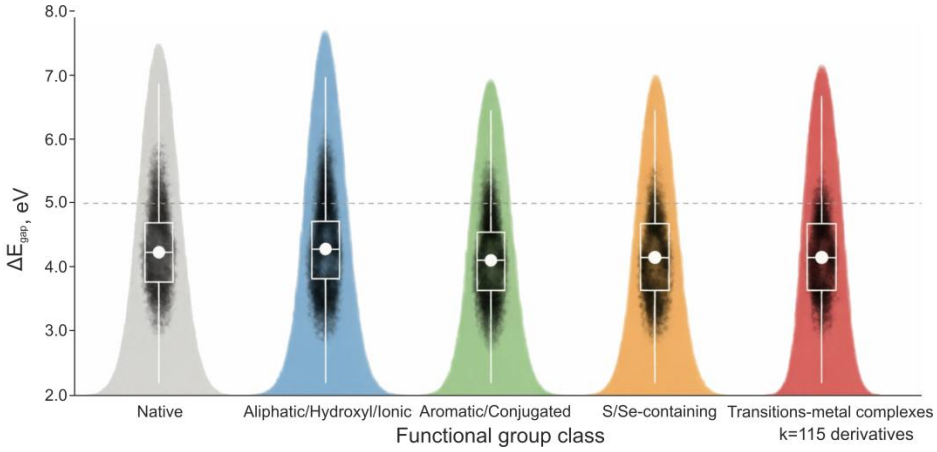


Fig. 2 Distribution of the HOMO–LUMO energy gap (ΔE_{gap}) across different classes of functional groups. The violin plots illustrate the probability density of the ΔE_{gap} values, while the embedded box plots show the median, interquartile range, and whiskers. The data clearly demonstrate a systematic reduction in the energy gap from aliphatic/hydroxyl derivatives (5.1–5.6 eV) to aromatic (3.7–4.8 eV), sulfur/selenium-containing (3.2–3.5 eV), and transition-metal-coordinated derivatives (2.6–2.8 eV), highlighting the profound impact of π-conjugation, heteroatom polarizability, and metal-ligand hybridization on intramolecular charge-transfer capability.

3.1.4. Electronic modulation via transition-metal coordination

The smallest HOMO–LUMO energy gaps within the investigated dataset are observed for transition-metal-coordinated derivatives, including cobalt-, vanadate-, and titanate-modified chitosan. Their ΔE_{gap} values collapse to the 2.6–2.8 eV range.

This dramatic narrowing is primarily associated with the participation of partially occupied metal d-orbitals. The hybridization between the metal-centered d-states and the coordinating amino/hydroxyl ligands of the polysaccharide backbone introduces additional electronic states in the vicinity of the frontier molecular orbitals. This metal-ligand orbital mixing substantially facilitates intramolecular charge transfer and maximizes the global electrophilicity index (ω). While these molecular-level characteristics do not directly equate to macroscopic solid-state electronic conductivity (which additionally requires efficient intermolecular orbital overlap), they confirm that transition-metal coordination is the most effective molecular strategy for transforming electronically rigid polysaccharides into highly reactive, electronically soft molecular architectures suitable for electrocatalysis and advanced sensor platforms.

To visually corroborate the calculated energetic trends and orbital localization patterns, the spatial distributions of the frontier molecular orbitals for representative chitosan derivatives are presented in Fig. 3. The isosurfaces clearly illustrate the progressive shift from localized orbitals on the glucosamine backbone in native and hydroxyalkylated chitosan to extensive π-delocalization in the seleno-derivative, culminating in pronounced metal-ligand orbital hybridization in the cobalt-coordinated complex.

3.1.5. Comparative analysis: Chitosan vs. Cellulose derivatives

Although both polysaccharides share a similar ß-(1→4) linked carbohydrate backbone, chitosan derivatives exhibit significantly broader electronic tunability than their cellulose counterparts. This difference is fundamentally attributed to the presence of the C-2 primary amino group in the chitosan repeating unit.

The nitrogen lone pair serves as a highly effective anchoring center and electron donor, allowing for stronger electronic coupling with aromatic substituents, sulfur-containing fragments, and transition metals. In contrast, cellulose lacks this amino functionality; its electronic modulation relies almost exclusively on the modification of hydroxyl groups. Consequently, cellulose derivatives generally retain larger HOMO–LUMO gaps (ΔE_{gap}>5.0 eV) unless heavily functionalized with extended aromatic systems (e.g., phthalate, trimellitate, or nucleoside-containing substituents), which can reduce the gap to the 3.3–4.4 eV range via π-conjugation and donor-acceptor push-pull mechanisms. Overall, chitosan provides superior electronic tunability at the molecular level, whereas cellulose offers higher intrinsic structural and electronic stability—a complementary pairing that can be strategically exploited in the design of hierarchical composite biomaterials.

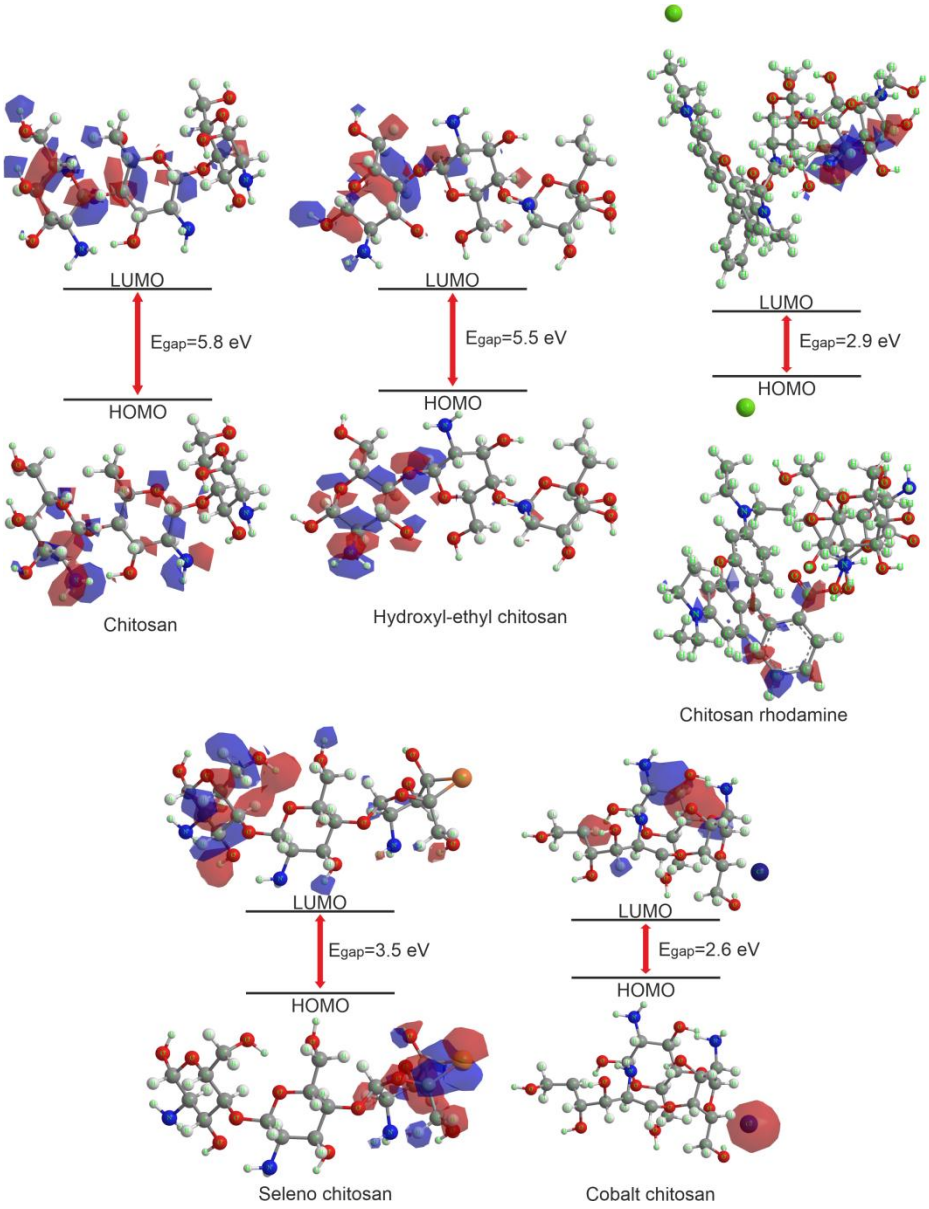


Fig. 3 Spatial distributions of frontier molecular orbitals for representative chitosan derivatives. HOMO (top row) and LUMO (bottom row) isosurfaces (isovalue = 0.02 a.u.) are shown for: (a) native chitosan; (b) hydroxyethyl-chitosan; (c) azo-chitosan; (d) seleno-chitosan; and (e) cobalt-chitosan complex.

3.2. Global conceptual DFT reactivity descriptors

The global conceptual DFT descriptors, derived from frontier molecular orbital energies via Koopmans' approximation, provide a rigorous quantitative framework for evaluating the chemical reactivity, kinetic stability, and intramolecular charge-transfer propensity of the investigated polysaccharide derivatives.

The ionization potential ($IP=-E_{HOMO}$) and electron affinity ($EA=-E_{LUMO}$) directly quantify the electron-donating and electron-accepting capabilities of the molecular systems, respectively. Derivatives functionalized with electron-rich moieties exhibit elevated EHOMO levels, consequently yielding lower IP values and enhanced nucleophilic character. Conversely, the introduction of strongly electron-withdrawing

groups (e.g., sulfonate, phosphate, or nitro-aromatics) significantly increases the EA, rendering these derivatives highly susceptible to nucleophilic attack and promising for applications in electrochemical sensing and redox catalysis.

The chemical hardness (η) serves as a fundamental indicator of molecular stability and resistance to external perturbation of the electron cloud. Derivatives possessing small η values—predominantly sulfur- and selenium-containing, aromatic, and transition-metal-coordinated systems—exhibit high molecular polarizability and electronic softness. These "soft" molecules are characterized by enhanced intramolecular charge-transfer capability and kinetic reactivity. In contrast, "hard" derivatives, such as hydroxyalkyl- and alkyl-substituted chains, maintain large ΔE_{gap} values and high η, reflecting their electronically rigid nature and superior kinetic stability. The global softness (S) exhibits the expected inverse correlation, further confirming that functionalization with highly polarizable heteroatoms or extended π-conjugated systems fundamentally softens the molecular electronic structure.

The electrophilicity index (ω) provides a comprehensive thermodynamic metric for the energy lowering due to maximum electron flow between a donor and an acceptor. Molecules that combine low chemical hardness with high electron affinity—most notably metal-coordinated complexes and heavily sulfated derivatives—exhibit the highest ω values. This indicates a strong thermodynamic driving force for electron acceptance, suggesting their favorable behavior in donor–acceptor intermolecular interactions, catalytic active sites, and charge-transfer biointerfaces.

When evaluated through the lens of global reactivity descriptors, the fundamental differences between chitosan and cellulose derivatives become quantitatively evident. Although both biopolymers share a similar ß-(1→4)-linked carbohydrate backbone, the presence of the C-2 primary amino group in chitosan provides a highly effective anchoring center and electron donor. This structural feature enables substantially greater modulation of the global reactivity indices (η, S, and ω) upon functionalization with aromatic, heterocyclic, or metal-coordinating substituents. Consequently, chitosan derivatives span a much broader reactivity spectrum, ranging from highly stable, hard molecular systems to highly reactive, soft electron-transfer agents. Cellulose derivatives, lacking this amino functionality, generally retain higher chemical hardness ($\eta > 2.5$ eV) and larger ΔE_{gap} values, unless heavily functionalized with strongly electron-withdrawing aromatic fragments (e.g., phthalate or trimellitate). Nevertheless, the aromatic cellulose derivatives that do achieve substantial electronic activation retain the superior mechanical stability and crystallinity inherent to the cellulose backbone. Ultimately, chitosan offers superior electronic and chemical tunability, whereas cellulose provides a more robust, kinetically stable scaffold—a complementary pairing that can be strategically exploited in the design of hierarchical composite biomaterials.

3.3. Exploration of the chemical space via unsupervised machine learning

3.3.1. Dimensionality reduction and Principal Component Analysis (PCA)

To elucidate the underlying structure–property relationships and visualize the multidimensional electronic chemical space of the 115 derivatives, Principal Component Analysis (PCA) was applied to the standardized conceptual DFT descriptor set.

Notably, the first two principal components (PC1 and PC2) accounted for 69.23% and 30.02% of the total variance, respectively, explaining 99.25% of the cumulative variance within the dataset. This exceptionally high explained variance is not coincidental but reflects the intrinsic mathematical collinearity of the global conceptual DFT descriptors. Since parameters such as chemical hardness, softness, chemical potential, and electrophilicity are algebraically derived from only two independent variables (E_{HOMO} and E_{LUMO}) via Koopmans' approximations, the PCA effectively reconstructs the fundamental two-dimensional electronic landscape defined by the frontier orbitals.

Loading vector analysis (Fig. 4) reveals the physical meaning of these components. The first principal component (PC1) represents a global reactivity and molecular softness gradient. It is strongly positively correlated with global softness and electrophilicity, and negatively correlated with the HOMO–LUMO energy gap and global hardness. Moving along the positive axis of PC1 corresponds to a transition from electronically soft, highly reactive molecules with narrow energy gaps to electronically rigid, kinetically stable systems with wide gaps. The second principal component (PC2) predominantly reflects the electron-donating capability, driven primarily by the ionization potential and the E_{HOMO} energy level, allowing for the discrimination between electron-rich and electron-deficient derivatives.

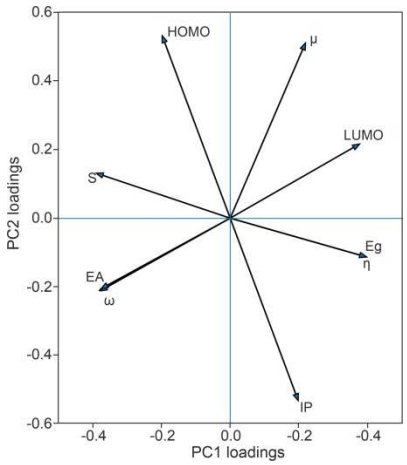


Fig. 4 PCA loading plot of the conceptual DFT descriptors. The loading vectors illustrate the contribution of E_{HOMO}, E_{LUMO}, ΔE_{gap}, IP, EA, μ, η, S, and ω to the first two principal components. PC1 is primarily associated with molecular reactivity descriptors (ΔE_{gap}, η, S, and ω), whereas PC2 mainly reflects electron-donating ability through HOMO energy and ionization potential.

3.3.2. Unsupervised clustering of polymer derivatives

To systematically classify the investigated derivatives according to their electronic profiles, K-Means clustering was performed. The optimal number of clusters was determined using the Elbow method and Silhouette analysis, as illustrated in Figure S2 (Supporting Information). Based on internal validation metrics (Silhouette score = 0.478, Davies–Bouldin index = 0.717, Calinski–Harabasz index = 140.99), the optimal solution partitioned the dataset into three distinct clusters (Table 2, Fig. 6). The spatial distribution of the cluster centroids in the PCA space (Table 3) confirms a clear separation of the derivatives into chemically meaningful groups:

Cluster 1 (Kinetically Stable, Electronically Rigid Derivatives): This is the largest group (44.3% of the dataset), located at positive PC1 values. It comprises standard hydroxyalkyl-, alkyl-, and simple amino-substituted derivatives characterized by wide HOMO–LUMO gaps (ΔE_{gap}≈5.366 eV) and high global hardness (η≈2.68 eV). According to the Maximum Hardness Principle, these molecules exhibit high kinetic stability, low molecular polarizability, and limited intramolecular charge-transfer capability, making them suitable for structurally robust, chemically inert biopolymer platforms.

Cluster 0 (Highly Reactive, Electronically Soft Derivatives): Located at strongly negative PC1 values, this cluster (27.0%) contains the most electronically intriguing molecules. It includes sulfur- and selenium-containing derivatives, extended aromatic conjugates, and transition-metal complexes. These systems exhibit significantly narrowed energy gaps (ΔE_{gap}≈3.70 eV), enhanced global softness (S=0.28 eV^{-1}), maximum electrophilicity (ω=3.41 eV). These characteristics indicate a high propensity for intramolecular charge redistribution and molecular recognition, highlighting their potential for electrochemical sensing and redox-active biointerfaces.

Cluster 2 (Electron-Deficient Derivatives): Separated primarily along the negative PC2 axis, this cluster (28.7%) consists of derivatives functionalized with strongly electron-withdrawing ionic groups (e.g., sulfonates, sulfates, and polycarboxylates). These molecules exhibit the lowest-lying HOMO levels (≈−6.21 eV) and the highest ionization potentials. The pronounced negative inductive effects strongly withdraw electron density from the polymer backbone, rendering these derivatives highly resistant to oxidation and ideal for applications requiring strong electrostatic interactions and ion-coordination capabilities.

Table 2 Cluster size distribution.

Cluster	Number of molecules	Percentage
Cluster 0	31	27.0%
Cluster 1	51	44.3%
Cluster 2	33	28.7%

Table 3 Cluster centroid coordinates in the PCA space.

Cluster	PC1	PC2
Cluster 0	−3.333	0.782
Cluster 1	2.039	0.814
Cluster 2	−0.021	−1.993

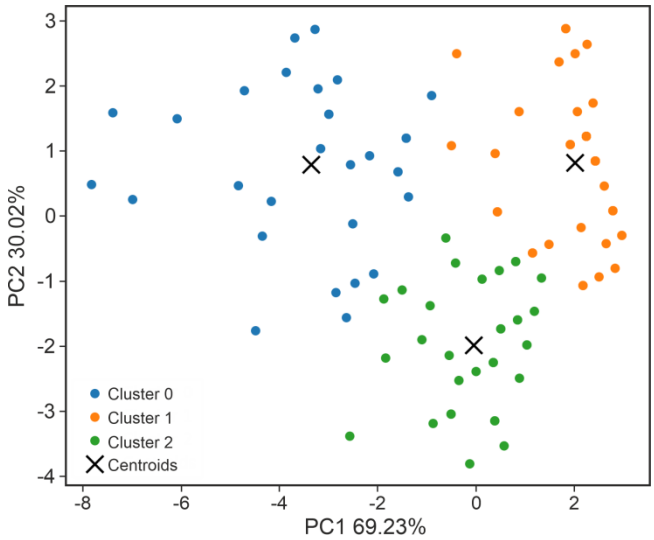


Fig. 5 Combined PCA score plot and K-Means clustering of 115 chitosan and cellulose derivatives. Points are color-coded by their designated K-Means clusters (k=3). Star markers highlight structural anomalies identified by the Isolation Forest algorithm, including metallocomplexes (chitosan-vanadate, cobalt-chitosan, chitosan-titanate) and dye conjugates (chitosan-rhodamine) exhibiting exceptionally narrow HOMO–LUMO energy gaps (ΔE_{gap}<3.0 eV).

3.3.3. Anomaly detection and identification of novel electronic scaffolds

In high-throughput computational screening, identifying statistical "outliers" is a powerful strategy for discovering molecular architectures with exceptional electronic properties. Utilizing the Isolation Forest algorithm, we isolated 5% of the dataset as anomalous structures that drastically deviate from the standard polysaccharide chemical space.

The anomaly detection predominantly flagged transition-metal complexes (specifically, chitosan-vanadate, cobalt-chitosan, and chitosan-titanate). The pronounced electronic deviations originate

from the participation of partially occupied transition-metal d-orbitals. The hybridization between the metal-centered d-states and the coordinating amino/hydroxyl ligands of the chitosan backbone introduces additional electronic states near the frontier orbitals. This metal-ligand orbital mixing leads to a dramatic collapse of the HOMO–LUMO energy gap down to 2.60–2.80 eV, accompanied by extreme electrophilicity (ω>5.0 eV). Such narrow molecular gaps indicate that the coordination of d-block metals endows the biopolymer architecture with exceptional electronic softness and intramolecular charge-transfer capability, characteristic of highly polarizable organometallic systems.

Furthermore, the chitosan-rhodamine conjugate was identified as a distinct organic outlier. Unlike the metal complexes, its anomalous electronic behavior arises from the heavy delocalization of π-electrons introduced by the dye system. The extended conjugated xanthene chromophore significantly increases the molecular polarizability and global softness, reducing the ΔE_{gap} primarily through extensive π-electron delocalization across a large molecular framework rather than through metal-ligand hybridization.

Overall, the unsupervised machine-learning pipeline provides a robust, data-driven framework for navigating the polysaccharide chemical space. It enables the rational prioritization of electronically soft derivatives for advanced charge-transfer applications and electronically rigid derivatives for stable biomaterial platforms, while successfully flagging unique organometallic and highly conjugated scaffolds for targeted experimental validation.

4 Conclusion

In this study, we established a comprehensive in silico screening framework for 115 chemically modified chitosan and cellulose derivatives. By integrating density functional theory (DFT) calculations with unsupervised chemoinformatics workflows, we systematically mapped the frontier molecular orbital (FMO) characteristics and global conceptual reactivity descriptors of these biopolymers to guide their rational molecular design.

Our quantum chemical analysis demonstrates that native polysaccharide backbones exhibit wide HOMO–LUMO energy gaps (ΔEgap≈5.8–6.1 eV), characteristic of electronically rigid and kinetically stable molecular systems with low intramolecular charge-transfer capability. However, rational structural modification enables the systematic tuning of their intramolecular electronic properties across a broad energetic range (> 3 eV). Specifically, the incorporation of aromatic π-conjugated systems, highly polarizable chalcogens (sulfur and selenium), and transition-metal coordination centers significantly narrows the ΔE_{gap} to the 2.6–4.5 eV range. Transition-metal-coordinated chitosan derivatives exhibit the most pronounced reduction (ΔEgap≈2.6–2.8 eV), driven by metal-ligand orbital hybridization, which imparts exceptional molecular softness and intramolecular charge-transfer capability.

Comparative structural analysis reveals that chitosan derivatives possess significantly broader electronic tunability than their cellulose counterparts. This difference is fundamentally governed by the C-2 primary amino group in chitosan, which acts as a highly effective electron-donating anchor for functionalization. In contrast, achieving comparable electronic activation in cellulose requires extensive aromatic substitution (e.g., phthalate or trimellitate fragments), reflecting its higher intrinsic structural and electronic rigidity.

From a chemoinformatics perspective, the integration of unsupervised machine-learning techniques—specifically Principal Component Analysis (PCA), K-Means clustering, and Isolation Forest anomaly detection—proved highly effective for navigating the high-dimensional descriptor space. This data-driven approach successfully partitioned the chemical space into distinct electronic clusters and identified transition-metal complexes and specific dye conjugates as statistical outliers with exceptional frontier orbital characteristics, thereby highlighting non-intuitive structure–property relationships.

Ultimately, the generated molecular electronic-property map provides a robust, data-driven foundation for the rational, computer-aided design of next-generation functional biopolymers for bioelectronics, electrochemical sensing, and redox-active biointerfaces. While this conceptual DFT study effectively captures intrinsic intramolecular charge-transfer capabilities, future computational work will extend this framework to periodic solid-state calculations to evaluate macroscopic charge transport and intermolecular packing effects in these advanced polysaccharide-based materials.

Declarations

Funding

This research was supported by the Basic Fundamental Program of the Academy of Sciences of the Republic of Uzbekistan and the fundamental project funded by the Ministry of Higher Education, Science and Innovation of the Republic of Uzbekistan (Grant No. FZ-4721055613).

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Ethics approval

Not applicable. This study did not involve human participants or animals.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

The datasets generated during and/or analyzed during the current study (including computed DFT descriptors and molecular structures) are available from the corresponding author on reasonable request.

Code availability

The custom Python scripts used for the unsupervised machine learning workflow (PCA, K-Means, Isolation Forest) are available from the corresponding author on reasonable request.

Authors' contributions

All authors contributed to the study conception and design. Material preparation, data collection and DFT calculations were performed by Nilufar Burkhanova and Ilnar Nurgaliev. The machine learning workflow and data analysis were performed by Ziyoda Abrarova and Akmal Abilkasimov. The first draft of the manuscript was written by Ilnar Nurgaliev and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

[1] Sana PA, Thanha KPK, Pramod K (2023) Recent advances in harnessing of natural polymers for bioelectronics. Int J Biol Macromol 242:124856. https://doi.org/10.1016/j.ijbiomac.2023.124856

[2] Javanbakht F, Saeb MR, Salami-Kalajahi M (2026) Renewable polymer electrolytes: Toward sustainable wearable energy devices. Adv Ind Eng Polym Res. https://doi.org/10.1016/j.aiepr.2026.06.005

[3] Tran VV (2026) Polymers-based biosensors and bioelectronics: Designs and applications. Biosensors 16:211. https://doi.org/10.3390/bios16040211

[4] Di Pasquale G, Pollicino A (2026) Conductive hydrogels for wearable sensing: Materials, mechanics, and multimodal interfaces. Appl Sci 16:7546. https://doi.org/10.3390/app16157546

[5] Robert B, Kumar SS, Thomas T, Kalpathy SK (2026) Can recycled e-waste polymers power the future of sensors? J Mater Chem A 14:13318–13331. https://doi.org/10.1039/d5ta09015f

[6] Ali NM, Hefnawy MA, Medany SS, Ismail KM (2026) Polymer-based capacitors for high-performance energy storage: Materials, mechanisms, and recent advances. J Alloys Compd 1062:187382. https://doi.org/10.1016/j.jallcom.2026.187382

[7] Shaji S, Anna A, Kiran S, Aboobaker A, Varghese AJ, Nancy P, Ravindran L, Thomas S (2023) Review on sustainable flexible electronics: Exploring the potential of chitosan, cellulose, starch, silk fibroin and gelatin. Discov Polym 1:19. https://doi.org/10.1007/s44347-023-00031-7

[8] Zhang Z, Li Y, Wang S (2022) Molecular engineering of cellulose-based materials for flexible

electronics and advanced applications. Adv Funct Mater 32:2106854. https://doi.org/10.1002/adfm.202106854

[9] Annu, Han SS, Shin DK (2023) Chitosan/MXene hydrogels for biomedical and self-powered health-monitoring bioelectronics: Challenges and strategic solutions. Int J Biol Macromol 253:127345. https://doi.org/10.1016/j.ijbiomac.2023.127345

[10] Rafati N, Zarepour A, Khosravi A, Iravani S, Zarrabi A (2023) Chitosan–cellulose hydrogels: Advances in stimuli-responsive biomedical therapeutics. RSC Adv 13:20575–20599. https://doi.org/10.1039/d3ra02743k

[11] Argilashki D, Uzunova Y (2026) Enhancing drug delivery through chemical modification of chitosan: A review. Pharmacia 73:1–14. https://doi.org/10.3897/pharmacia.73.e180564

[12] Zhao Y, Wang M, Li H, Zhang Z, Li H (2026) Advances in modification strategies of polysaccharides and their applications. J Food Sci 91:e70779. https://doi.org/10.1111/1750-3841.70779

[13] Pispas I, Papagiannopoulos A (2026) Next-generation polysaccharide-based nanocarriers for precision medicine: Structure–property principles, responsiveness, and therapeutic translation. Macromol 6:19. https://doi.org/10.3390/macromol6010019

[14] Liu YR, Zhan Y, Liu J, Sun C, Tang XG (2026) Research progress on drug delivery systems based on traditional chinese medicine polysaccharides for the treatment of colon cancer. J Mater Chem B 14:1494–1510. https://doi.org/10.1039/d5tb02175h

[15] Anand MAV, Yang C-J, Wang W-L, Yao C-H, Lai J-Y (2026) Marine-derived polysaccharides enhance the drug delivery mechanism in the management of ocular diseases: A review. Int J Biol Macromol 340:150199. https://doi.org/10.1016/j.ijbiomac.2026.150199

[16] Song R, Tang Z, Liu B, et al (2026) Structural characterization and prebiotic activity of polysaccharides from Suaeda glauca. Chem Biol Technol Agric 13:68. https://doi.org/10.1186/s40538-026-00965-5

[17] Uto T (2025) Atomistic simulations of polysaccharide materials for insights into their crystal structure, nanostructure, and dissolution mechanism. Polym J 57:33–41. https://doi.org/10.1038/s41428-024-00966-x

[18] Zhang C, Chen H, Yu G, Lai Y, Zou Y, Feng Z, Yao M, Chen Q (2026) Integrated in silico design of self-assembled herbal bioactives–polysaccharides nanoformulation for targeted therapy against gastric pathogen Helicobacter pylori. Carbohydr Polym 381:125184. https://doi.org/10.1016/j.carbpol.2026.125184

[19] Rusli RA, Abidin SZZ, Hussain NFS, Balian SRC (2026) Electrical, structural, and electronic insights of polysaccharide-based biopolymer electrolytes incorporated with sodium perchlorate salt. J Polym Mater 43:22. https://doi.org/10.32604/jpm.2026.077915

[20] Nurgaliev I, Marasulov M, Normuradov N, Usmanova M, Ashurov N (2026) Molecular modeling of gelatin–plasticizer interactions: Insights from DFT and molecular dynamics. Int J Comput Mater Sci Eng. https://doi.org/10.1142/S2047684126500041

[21] Fukui K (1982) Role of frontier orbitals in chemical reactions. Science 218:747–754. https://doi.org/10.1126/science.7123245

[22] Parr RG, Pearson RG (1983) Absolute hardness: Companion parameter to absolute electronegativity. J Am Chem Soc 105:7512–7516. https://doi.org/10.1021/ja00364a005

[23] Toropova AP, Toropov AA, Selvestrel G, Iovine N, Roncaglioni A, Benfenati E (2026) QSPR models for water solubility of organic compounds using correlation intensity index and Las Vegas algorithm. SAR QSAR Environ Res 37:299–310. https://doi.org/10.1080/1062936X.2026.2659325

[24] Zhao X, Yan Z, Zeng L, Zheng Y, Rong C (2026) Information-theoretic and conceptual density functional theory insights on frustration in molecular clusters. Entropy 28:213. https://doi.org/10.3390/e28020213

[25] Xie C, Qiu H, Liu L, You Y, Li H, Li Y, Sun Z, Lin J, An L (2025) Machine learning approaches in polymer science: Progress and fundamental for a new paradigm. SmartMat 6:e1320. https://doi.org/10.1002/smm2.1320

[26] Lee JM, Gao X, Yeong WY (2026) Physicochemical-informed predictive modelling on small datasets for designing conductive polymer inks in soft bioelectronics. npj Flex Electron 10:89. https://doi.org/10.1038/s41528-026-00587-9

[27] Albariqi AH, Alsalhi A, Alshamrani M, et al (2026) Predictive modeling of controlled drug release from polysaccharide-based systems using gradient boosting and metaheuristic optimization. Sci Rep 16:18769. https://doi.org/10.1038/s41598-026-48281-0

[28] Nurgaliev IN, Marasulov MB, Hamzayev AI, Abilkasimov AB (2026) Machine learning and AI in polymer science: A review of progress, challenges, and future directions. Uzbek J Polym 5:5–41. https://doi.org/10.66640/UJP-2026-5-00001

[29] Cao Z, Chen T, Xie J, Xie J (2026) Artificial intelligence in functional polysaccharides for food applications: Process optimization, structure–function decoding, and rational design. J Agric Food Chem 74:11954–11968. https://doi.org/10.1021/acs.jafc.6c00806

[30] Hemmami H, Ben Amor I, Ben Amor A, Zeghoud S, Ahmed S, Alhamad AA (2024) Chitosan, its derivatives, sources, preparation methods, and applications: A review. J Turk Chem Soc Sect A Chem 11:341–364. https://doi.org/10.18596/jotcsa.1336313

[31] Zia KM (2025) Preparation and functionalization of the chitosan derivatives. In: Chitosan. Smart

Nanomaterials Technology. Springer, Singapore. https://doi.org/10.1007/978-981-96-2629-8_2

[32] Morais A, Lima R, Pinto MMM, Tiritan ME, Fernandes C (2026) Chitosan derivatives: Challenges and opportunities in the green and sustainable transition era. Molecules 31:1273. https://doi.org/10.3390/molecules31081273

[33] Žigrayová D, Mikušová V, Mikuš P (2024) Advances in chitosan derivatives: Preparation, properties and applications in pharmacy and medicine. Gels 10:701. https://doi.org/10.3390/gels10110701

[34] Garg T, Arora S, Pahwa R (2025) Cellulose and its derivatives: Structure, modification, and application in controlled drug delivery. Futur J Pharm Sci 11:76. https://doi.org/10.1186/s43094-025-00834-2

[35] El-Saadony MT, Saad AM, Alkafaas SS, Dladla M, Ghosh S, Elkafas SS, et al (2025) Chitosan, derivatives, and its nanoparticles: Preparation, physicochemical properties, biological activities, and biomedical applications – A comprehensive review. Int J Biol Macromol 313:142832. https://doi.org/10.1016/j.ijbiomac.2025.142832

[36] Celada LM, Olsén P (2026) Cellulose at scale: From past challenges to future opportunities for sustainable functional materials. J Mater Chem A 14:28149–28165. https://doi.org/10.1039/d6ta02233b

[37] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, et al (2016) Gaussian 16, Revision C.01. Gaussian, Inc., Wallingford, CT

[38] Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1994) Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. J Phys Chem 98:11623–11627. https://doi.org/10.1021/j100096a001

[39] Francl MM, Pietro WJ, Hehre WJ, Binkley JS, Gordon MS, DeFrees DJ, Pople JA (1982) Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. J Chem Phys 77:3654–3665. https://doi.org/10.1063/1.444267

[40] Weigend F, Ahlrichs R (2005) Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. Phys Chem Chem Phys 7:3297–3305. https://doi.org/10.1039/b508541a

[41] Marenich AV, Cramer CJ, Truhlar DG (2009) Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. J Phys Chem B 113:6378–6396. https://doi.org/10.1021/jp810292n

[42] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al (2011) Scikit-learn: Machine learning in Python. J Mach Learn Res 12:2825–2830.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SIArticleNurgaliev.docx