

Machine learning-driven SSbD strategy for discovering next-generation anodes in solid oxide electrolysis cells

Dawid Falkowski^{a,b}, Alicja Mikolajczyk^{a,b,*}, Piotr Skurski^{a,b}, Jakub Brzeski^{a,b},
Tomasz Puzyn^{a,b,**}

^a Faculty of Chemistry, University of Gdansk, Gdansk, Poland

^b QSAR Lab Ltd, Gdansk, Poland

ARTICLE INFO

Keywords:

Solid oxide electrolysis cell
Solid oxide fuel cell
Safe and sustainable by design
Artificial intelligence
Machine learning
Density functional theory

ABSTRACT

Designing robust and efficient anode materials remains a key challenge in advancing Solid Oxide Electrolysis Cell (SOEC) technology for sustainable hydrogen production. This study presents a computational framework based on the Safe and Sustainable by Design (SSbD) concept, integrating density functional theory (DFT) with artificial intelligence (AI)-driven analysis to accelerate the discovery of Brownmillerite-type oxides for SOEC anodes. A novel site-specific nomenclature enables systematic screening of 240 doped structures with diverse A-site and B-site substitutions. Ground-state spin multiplicities are determined, and over 6000 DFT calculations are performed to extract electronic and surface descriptors. These data are incorporated into an unsupervised machine learning (ML) pipeline using principal component analysis (PCA) and k-means clustering to reveal structure–property relationships and classify materials by functional potential. Medoid-based selection identifies representative candidates from each cluster. Key descriptors – including polar/nonpolar surface areas, electrostatic potential variance, and internal charge separation – are correlated with performance indicators such as water adsorption, surface stability, and redox balance under SOEC operating conditions. The framework demonstrates a scalable, AI-augmented SSbD approach for materials discovery in electrochemical energy systems and establishes a rational basis for experimental validation of next-generation SOEC anodes.

1. Introduction

Solid Oxide Fuel Cells (SOFCs) represent one of the most promising Key Enabling Technologies for a sustainable hydrogen economy [1]. Their dynamic development in recent years has been driven not only by technological advancements but also by policy initiatives aimed at accelerating the energy transition, particularly within the European Union under frameworks such as the European Green Deal and REPowerEU plans [2–4]. These high-efficiency devices are capable of supplying both industrial facilities and residential buildings with low-emission energy, supporting the global shift away from fossil fuels.

A particularly transformative feature of SOFC systems is their reversibility. When operated in electrolysis mode, they function as Solid Oxide Electrolysis Cells (SOECs) – also referred to as reversible Solid Oxide Cells (rSOCs) – enabling a closed-loop energy system based on the Power-to-Hydrogen-to-Power (P2H2P) concept [5]. In this

configuration, SOECs convert surplus renewable electricity into hydrogen through high-temperature steam electrolysis, while SOFCs recover that stored hydrogen later to generate electricity on demand (Fig. 1). This bidirectional functionality supports grid balancing and energy storage and aligns well with decarbonization strategies targeting renewable integration and hydrogen-based energy carriers [6]. At the same time, the practical viability of SOECs depends directly on the resilience of materials to operating conditions: high temperature, high steam partial pressure, and transient redox states associated with start-up, shut-down, and load-following [7,8].

SOEC technology offers high electrical efficiency and scalability, making it suitable for a wide range of applications from megawatt-scale industrial systems to small-scale residential units [9,10]. However, a key challenge in deploying SOECs broadly lies in optimizing their component materials to meet application-specific performance and durability requirements. For example, an SOEC continuously operated in an

This article is part of a special issue entitled: ANM2025 (Titus) published in International Journal of Hydrogen Energy.

* Corresponding author. Faculty of Chemistry, University of Gdansk, Gdansk, Poland.

** Corresponding author. Faculty of Chemistry, University of Gdansk, Gdansk, Poland.

E-mail addresses: a.mikolajczyk@qsarlab.com (A. Mikolajczyk), t.puzyn@qsarlab.com (T. Puzyn).

<https://doi.org/10.1016/j.ijhydene.2026.154543>

Received 27 November 2025; Received in revised form 28 February 2026; Accepted 13 March 2026
0360-3199/© 2026 Published by Elsevier Ltd on behalf of Hydrogen Energy Publications LLC.

industrial setting faces significantly different stressors than one used intermittently in a household context [11]. These differences have specific physicochemical origins. Steam-rich and redox environments promote microstructural reconstruction and the loss of active reaction sites, may induce volume changes and mechanical damage, and can trigger interfacial reactions and secondary phase formation at electrode-electrolyte boundaries. In parallel, they modify defect chemistry, which collectively translates into evolving polarization losses and faradaic efficiency over time [7,12].

Fundamentally, SOECs consist of a solid oxide electrolyte, an anode, a cathode, and interconnects. In electrolysis mode, steam is reduced at the cathode to produce hydrogen, while in fuel cell mode, hydrogen is oxidized at the anode to generate electricity [13]. Among candidate materials, perovskite-derived structures – particularly Brownmillerite-type oxides – have emerged as leading contenders for anode components due to their favorable electronic conductivity, redox stability, and structural flexibility [14]. In materials with ordered oxygen vacancies (such as brownmillerites), vacancy channels and compositional tunability provide a natural design lever to balance electronic transport, catalytic activity, and redox tolerance, which is particularly relevant in steam-rich and cycling conditions typical of SOEC operation [8]. Despite this promise, many of the materials that could enable efficient hydrogen technologies remain underexplored or are still in the early stages of development [15,16]. According to the *Materials 2030 Manifesto*, such systems fall under the umbrella of “Innovative Advanced Materials,” and their development benefits greatly from advanced computational design methodologies [17].

In practice, recent data-driven studies in adjacent SOC-related materials problems have mostly concentrated on either broad perovskite property classification or on electrolyte-centric transport targets rather than on the fuel-electrode-specific multi-objective constraints relevant to SOEC operation. Representative examples include ML workflows for predicting ionic conductivity in solid-state electrolytes [18], ML-assisted modeling of oxygen ionic conductivity across doped ABO_3 perovskites [19], and classification or surrogate prediction tasks for ABO_3 crystal properties and phonon-related descriptors [20,21]. Importantly for vacancy-ordered perovskite derivatives, neural-network interatomic potentials combined with molecular dynamics have been used to investigate oxide-ion transport in Co-doped brownmillerites [22], demonstrating that advanced ML can provide mechanistic transport insight when coupled to physics-based simulations. However, these advances typically address isolated property targets (e.g., ionic conductivity or diffusion-related metrics) and are rarely formulated as transparent down-selection pipelines that integrate stability, redox

robustness, defect chemistry, and operating-environment stressors into a single, interpretable screening logic for SOEC anode materials. This limitation is particularly relevant because diffusion coefficients and transport-related descriptors, while important for cell efficiency, are only one part of the coupled performance-durability landscape governing SOEC deployment at scale [9,10]. More broadly, the emerging paradigm of digital materials design explicitly encourages coupling multiple computational techniques (DFT, molecular dynamics, Monte Carlo methods, multiscale multiphysics modeling, numerical optimization, and ML) into end-to-end in silico workflows [23–25].

In response to this need, we present a proof-of-concept computational framework for the rational design of novel SOEC anode materials, integrating density functional theory (DFT) with virtual screening and unsupervised machine learning techniques (principal component analysis and k-means clustering). Throughout the manuscript, machine learning-driven refers to the use of unsupervised learning for data structuring and down-selection rather than supervised prediction. The framework is motivated by the Safe and Sustainable by Design (SSbD) paradigm [26], which emphasizes considering performance together with safety and sustainability aspects from the earliest stages of innovation. Recent SSbD developments in the EU emphasize a tiered and iterative approach, in which early screening stages support prioritization and reduce experimental workload before later, more comprehensive assessment stages are applied [27,28]. Importantly, this workflow provides dual value: it enables efficient, interpretable down-selection from a large compositional space to reduce experimental workload, while simultaneously supporting early-stage SSbD decision-making by keeping the screening criteria transparent and traceable for subsequent sustainability-focused assessment. Through this method, we aim to support informed decision-making in materials selection and to identify a subset of high-potential compounds for experimental validation. We emphasize that this is not a predictive machine learning model and that comprehensive SSbD evaluation requires additional data and dedicated methodologies beyond the scope of the present work. In line with the tiered and iterative nature of SSbD implementation, the workflow is intended as an early screening step that reduces experimental workload and helps focus later-stage more comprehensive SSbD assessments. By leveraging the synergy of quantum chemistry and AI-guided analysis, this work contributes a scalable tool for accelerating the discovery of next-generation anode materials for solid oxide electrolysis applications.

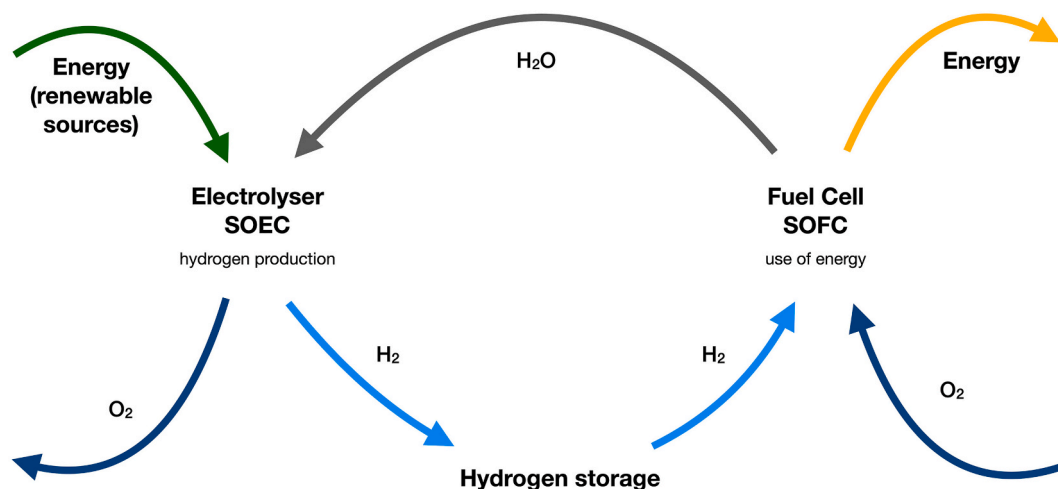


Fig. 1. Schematic representation of a closed-loop energy system utilizing SOEC for hydrogen production and SOFC for energy conversion. The system integrates renewable energy sources for electrolysis and hydrogen storage for later energy use.

2. Methods

2.1. Simplified nomenclature for brownmillerites structures

In our work, we propose a new, simplified nomenclature system for mixed oxides with a Brownmillerite-type structure. Traditional nomenclature, such as $\text{Ca}_5\text{Sr}_4\text{Fe}_1\text{Co}_3\text{O}_{20}$, is not easily readable and is challenging for rapid interpretation in chemoinformatic analyses. The developed system, e.g., $\text{Sr}_4\text{Ca}_1\text{Co}_3\text{Fe}_1$, provides a more intuitive representation of the composition while retaining key structural information about the compound. The Brownmillerite structure is characterized by the presence of two distinct coordination centers: tetrahedral and octahedral, also referred to as the A-site and B-site, respectively. Introducing an underscore (“_”) in our notation explicitly separates elements occupying these two different sites, allowing for a quick understanding of their spatial organization. The number of oxygens was skipped. An example is shown in Fig. 2. General name pattern is presented below:

$$A(5-i)D_1(i) \dots B(4-j)D_2(j)$$

A – large cation (A-site), typically 7-coordinated in brownmillerite (e.g., Ca, La, Sr, Bi)

B – transition metal centers (B-site), subdivided into 6-coordinated (octahedral) and 4-coordinated (tetrahedral) positions, e.g., Fe, Co, Mn.

D_1 – dopant in tetrahedral center (calcium position)

D_2 – dopant in octahedral center (iron position)

i – number of dopant atoms, replaced tetrahedral center atoms

j – number of dopant atoms, replaced octahedral center atoms.

On the other hand, the use of LSM-type nomenclature as an abbreviation of Lanthanum Strontium Manganite or $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ [29] is inadequate in the context of chemoinformatic analyses, as it does not account for precise atomic ratios or their location in different coordination centers. This is a significant issue in materials research, where the distribution of cations affects their physicochemical properties.

As a result of applying custom nomenclature, the original names were shortened. The original compound names, listed alongside the new nomenclature, are presented in Supplementary Materials (Table S1).

2.2. Quantum mechanical calculations

Electronic energies of selected molecular structures were calculated using the density functional theory (DFT) method (Minnesota15 (MN15) functional [30]) with the Los Alamos National Laboratory (LANL) effective core potentials (ECP) with the appropriate valence basis set of

double zeta quality (denoted LANL2DZ [31,32]) (for Fe, Co, Mn, La, Sr, and Bi) and the 6-31G basis set (for Ca and O). Consistently, all molecular descriptors characterizing the studied systems were determined based on the electron density obtained at the MN15/LANL2DZ/6-31G level. The calculations were performed using Gaussian16 [33] computational chemistry software package and MultiWFN [34] multifunctional wavefunction analyzer.

2.3. Selection of a representative molecular fragment for electronic structure calculations

It is well known that Brownmillerite crystal structure has sheets of corner-sharing Fe–O octahedra connected to single chains of Fe–O tetrahedra whereas the calcium ions are surrounded by irregular arrays composed of seven oxygen ions each. Since quantum chemical electronic structure calculations for such large systems are impossible to perform, we had to reduce the size of the Brownmillerite structure. However, we did so in a manner that the fragments used in the calculations remained representative and pertained the crucial structural features of the investigated crystals. Namely, we generated two fragments (labeled M and S, as shown in Fig. 3).

Our test calculations for selected M and S structures showed that achieving convergence of the SCF procedure (within the DFT calculation) using the chosen basis set typically takes between 6 and 15 h of runtime for M fragments, whereas the analogous time was significantly shorter (2–5 h) for S fragments. These times represent the elapsed times for parallel calculations using 24 CPU cores and 280 GB RAM. Considering that we intended to analyze a large number of systems (due to the need to account for various dopants), and that the final calculations had to be preceded by determining the ground state multiplicity for each system, we decided to select the S fragment. To justify our decision to significantly reduce the structure of the studied fragment, it is worth noting here that even for the seemingly smaller S structure, this meant performing over 6000 DFT calculations.

In addition to the shorter wall time per calculation, the use of the S fragment was dictated by the feasibility of generating a uniform and sufficiently large dataset for screening. Moving to a larger M fragment would not only increase the cost of each DFT job, but would also substantially expand the number of distinct substitution patterns required to maintain full coverage of the dopant space, rendering the study impractical within a reasonable timeframe. Moreover, our exploratory M-fragment runs revealed serious SCF and geometry convergence difficulties, which often prevented successful completion and would preclude a systematic, like-for-like comparison between M and S models

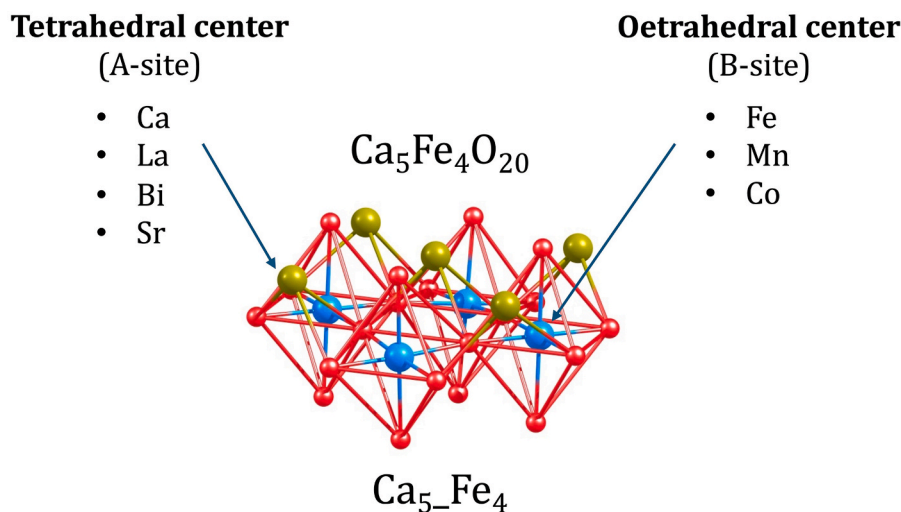


Fig. 2. Graphical explanation of simplified nomenclature for brownmillerites structures (in this case for $\text{Ca}_5\text{Fe}_4\text{O}_{20}$ structure). Tetrahedral and octahedral centers are distinguished.

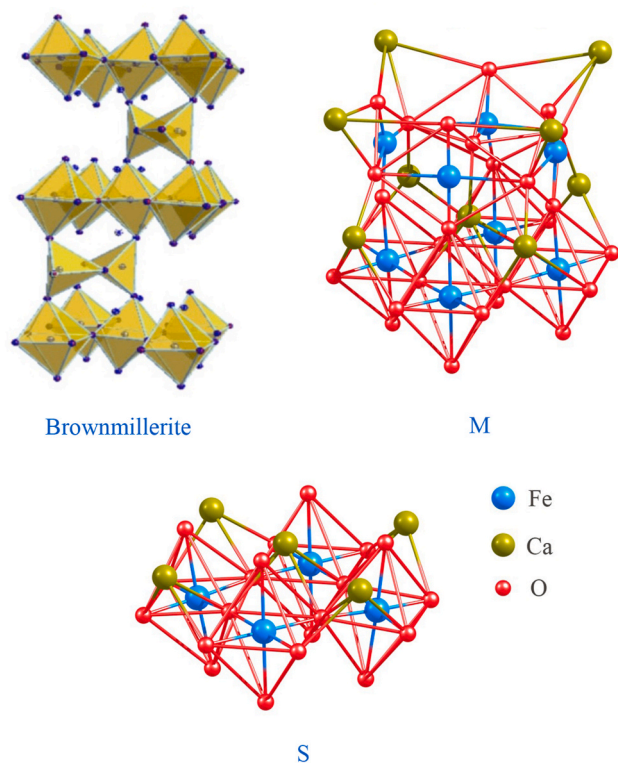


Fig. 3. Schematic Brownmillerite crystal structure and two fragments (M and S) selected for calculations.

based on a consistent set of computed properties. Therefore, the S fragment is adopted as the smallest tractable model that preserves the local coordination motifs relevant to the A-site and B-site substitutions studied here, enabling an internally consistent, high-throughput descriptor-based analysis.

Selecting a portion of the Brownmillerite structure labeled as S enabled to assess the impact of substituting Fe atoms with Co and/or Mn atoms, as well as replacing Ca atoms with La, Sr or Bi atoms, on the system's properties, we opted for. This choice allows us to consider all possible combinations of dopants. Namely, for the substitution of four iron atoms with Co and Mn dopants, there are 15 possible combinations. Similarly, in the case of replacing five Ca atoms in the S structure with dopants of the same type (La, Sr or Bi), there are 75 possible combinations in these three scenarios. In total, this results in 240 S-type systems, differing in the content of the mentioned metals. Consequently, it enables the determination of selected physicochemical properties based on the stoichiometry (molar fractions) indicating the percentage content of individual components. Since the selection of specific Fe or Ca atoms to be replaced by Mn and Co dopants or La, Sr and Bi dopants, respectively, was arbitrary, our strategy involved choosing atom locations where substitution with a dopant would result in the highest possible symmetry point group.

It is important to emphasize that the formulas presented in this study do not correspond to crystallographic unit cells, but rather to molecular fragments extracted from the brownmillerite structure and used in our quantum chemical calculations. These stoichiometries reflect the atomic content of the selected finite models and were chosen to preserve key coordination motifs and electronic characteristics of the parent materials.

The present computational protocol employs finite molecular fragments treated with DFT method to enable high throughput exploration of a large dopant space and to extract consistent local electronic and bonding descriptors across all compositions. As a consequence, the model is not designed to reproduce key bulk, extended, or transport

properties of brownmillerite oxides that are critical for SOEC anode performance. In particular, oxygen vacancy formation energies and equilibrium non-stoichiometry, oxygen vacancy migration barriers (and thus ionic transport), and the electronic band structure (including band dispersion, carrier effective masses, and long-range carrier transport) are outside the scope of this fragment representation. Additionally, long-range periodic electrostatics, collective lattice relaxations, and correlation effects that may require periodic approaches are not explicitly addressed here. Therefore, the descriptors and the PCA plus clustering results should be interpreted as a comparative screening and prioritization within a fixed modeling protocol, intended to identify chemically diverse, promising candidates for subsequent SOEC level validation using periodic calculations (including vacancy energetics and migration) and, ultimately, experimental testing.

2.4. Combined PCA and k-means approach

Using molecular descriptors derived from the quantum mechanical calculations conducted previously, we performed virtual screening using Principal Component Analysis (PCA) [35] combined with k-means clustering algorithm [36] to determine similarities between compounds and select representative ones for further experimental evaluation. PCA was used for dimensionality reduction by transforming the original descriptor set into a smaller number of orthogonal principal components while preserving the dominant variance in the dataset. During the PCA analysis, the number of retained principal components was selected using Cattell's scree test based on the scree plot. Descriptors were considered relevant for interpretation based on their factor loadings. Variables with an absolute loading value equal to or greater than 0.7 were treated as significant contributors to a given principal component. In parallel, based on the same descriptor set, the k-means clustering was performed. K-means clustering was performed on the same descriptor set and optimized using the standard inertia objective, defined as the within-cluster sum of squared Euclidean distances to the cluster centroids (squared Euclidean metric). The optimal number of clusters was selected using the elbow method applied to the inertia curve. Each cluster was assigned a distinct color for visualization compounds on the PCA-derived map were color-coded according to the k-means assignments. All calculations were performed in custom Python script.

2.5. Cluster composition analysis

The composition of individual clusters obtained through k-means analysis was examined using histograms, which depict the distribution of specific atoms within the clusters. The data was binned and visualized to capture trends in atomic compositions of the clustered compounds. For each cluster a set of histograms was prepared. On the left side, the distribution of individual elements occupying tetrahedral positions is presented, while on the right side, the elements located in octahedral centers are shown. This representation allows for the assessment of elemental variability within each cluster and the identification of patterns and similarities among the compounds.

2.6. Selection of representatives of clusters

Finally, using the medoid method [37], we selected the most representative samples for each cluster were chosen. This method was chosen because, unlike the centroid method, the representative point is an actual data point rather than a theoretical point that does not correspond to any actual data point in real existing data. The selection of the medoid is based on minimizing the sum of distances to all other points within the cluster.

3. Results and discussion

3.1. Determining ground state spin multiplicities

Each of the 240 compounds considered is an open-shell system, capable of being represented by electronic states of different multiplicities ($2S + 1$). Since it is not possible to determine the optimal multiplicity a priori (i.e., the one that minimizes the system's total energy), it was necessary to calculate the electronic energy for a range of $2S + 1$ values in all cases (odd $2S + 1$ values ranging from 1 to 39 for systems with an even number of electrons, and even $2S + 1$ values from 2 to 40 for systems with an odd number of electrons) and identify the preferred multiplicity for each specific case. By proceeding in this manner for each system (requiring the calculation of electronic energy for a total of 4800 electronic states), we determined the multiplicity of the ground electronic state for each of the 240 compounds. Further calculations (especially molecular descriptor calculations) were performed only for the ground electronic states. An exemplary dependence of electronic energy on multiplicity for one of the systems is illustrated in Fig. 4.

As indicated by our calculations, the lowest energy electronic states typically corresponded to relatively high values of spin multiplicity. Specifically, multiplicities in the range of 25–28 were the most represented (69 cases out of 240), whereas low values of $2S + 1$ (i.e., in the range of 1–4) corresponded to the lowest energy states in only 5 cases (again, out of 240). At the same time, we would like to emphasize that no far-reaching conclusions should be drawn from these findings, as in systems with such highly complex electronic structures (due to the presence of many heavy atoms), it is impossible to determine the influence of all factors on which state corresponds to the lowest energy state in a given system. Therefore, the sole outcome of the calculations described in this section was the determination of the spin multiplicity corresponding to the ground electronic states of all 240 molecular fragments considered (see Supplementary Materials, Table S1).

3.2. Descriptors development and data analysis

Using the developed quantum-chemistry-based workflow, we computed over 6000 DFT calculations to extract 48 electronic, structural, and surface descriptors for 240 doped Brownmillerite structures. Each compound's ground-state spin multiplicity was determined through exhaustive comparison of relative electronic energies across a

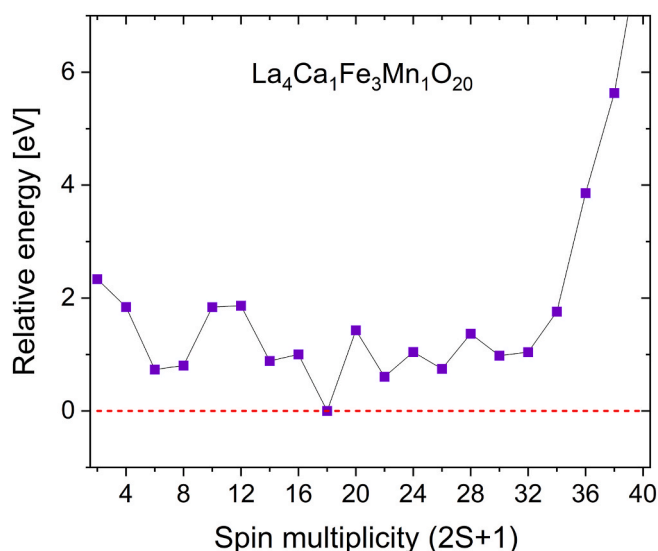


Fig. 4. Relative electronic energies of the representative $\text{La}_4\text{Ca}_1\text{Fe}_3\text{Mn}_1\text{O}_{20}$ system corresponding to various multiplicities (in this case, $2S + 1 = 17$ multiplicity corresponds to the ground electronic state).

wide multiplicity range, ensuring that all further descriptor calculations reflected the true lowest-energy state.

3.3. Dimensionality reduction via principal component analysis

To identify latent structure–property relationships within the dataset of 240 doped Brownmillerite compounds, we employed Principal Component Analysis (PCA) on a refined set of quantum-chemically derived descriptors. After removing near-constant and redundant features, PCA reduced the original 48-dimensional space to two principal components (PC1 and PC2), chosen based on the Cattell's criterion (Fig. 6), which together explained 63.44% of the total variance (PC1: 49.56%, PC2: 13.88%). PC1 was primarily associated with electronic descriptors, while PC2 captured structural averages and energy-related features (Fig. 7). PC1 predominantly encoded electronic descriptors, including polar and nonpolar surface areas, molecular polarity index (MPI), internal charge separation (ICS), electrostatic variance, dipole moment, and frontier orbital energies. PC2 was primarily linked to average structural energies and volumetric features. The interpretation of these components was guided by the normalized loadings (Fig. 5) and further elaborated in Section 3.6.

3.4. Clustering and representative compound selection

To uncover structurally and functionally distinct groups among the compounds, we applied k-means clustering to the PCA-reduced space. The optimal number of clusters was determined via the elbow method (Fig. 8), which indicated a clear inflection point at $k = 5$, beyond which inertia reduction plateaued. Clustering was thus performed in the two-dimensional PC1–PC2 space to ensure that grouping was informed by the most significant molecular features.

Each resulting cluster was visualized on the PCA map (Fig. 10), where the spatial proximity of points reflects descriptor-based similarity. To identify prototypical representatives for each cluster, we adopted a medoid-based strategy, selecting real, chemically feasible compounds located closest to the cluster centroids. This ensured that representative structures were not merely theoretical averages, but concrete candidates suitable for experimental synthesis and testing (Table 2).

3.5. Structural trends and dopant-driven cluster composition

Elemental composition within each cluster was examined to identify systematic trends. Histograms (Fig. 9, Table 1) illustrate the dopant distributions at tetrahedral (A-site) and octahedral (B-site) positions across clusters. Notably, the primary differentiator among clusters was the A-site doping pattern.

- **Cluster 1** ($n = 62$): Ca-rich systems with balanced La and Bi doping.
- **Cluster 2** ($n = 29$): La-dominant doping with Ca and negligible Bi/Sr.
- **Cluster 3** ($n = 43$): Sr-doped Ca-based systems, limited La/Bi.
- **Cluster 4** ($n = 30$): High Bi content and reduced Ca presence.
- **Cluster 5** ($n = 76$): Mixed Ca–La–Bi–Sr systems, with Ca predominance.

Importantly, B-site substitution (Fe/Mn/Co) did not significantly influence clustering, highlighting the dominant role of A-site dopants in shaping structure–property landscapes.

In more details, Group 1 (brown) consists of 62 compounds exhibiting a classic brownmillerite structure, primarily composed of systems where Ca occupies the tetrahedral center, and limited amounts of all other dopants (La, Bi, Sr) and mixture of Fe, Mn and Co in the octahedral center (Fig. 9a). Group 2 (turquoise) consists of 29 compounds, primarily composed of systems where Ca occupies the tetrahedral center, doped with La, and lack of Bi-, Sr-dopants. The group 2 consist of mixture of Fe, Mn and Co in the octahedral center (Fig. 9b). Group 3

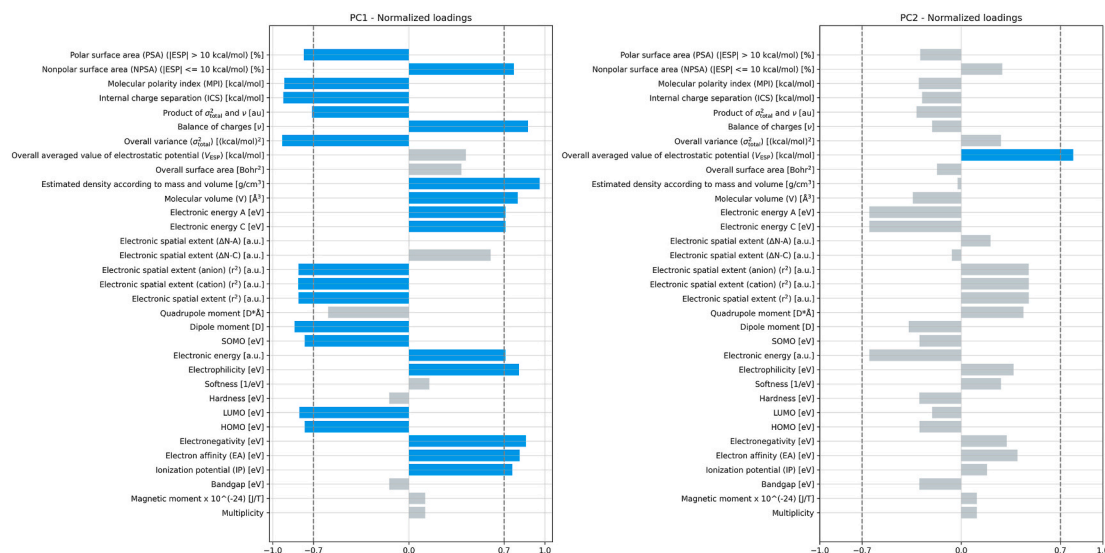


Fig. 5. Normalized factor loadings for the first two principal components (PC1 and PC2). Descriptors with significant contributions, as per Malinowski's criterion ($|\text{loading}| \geq 0.7$), are highlighted in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

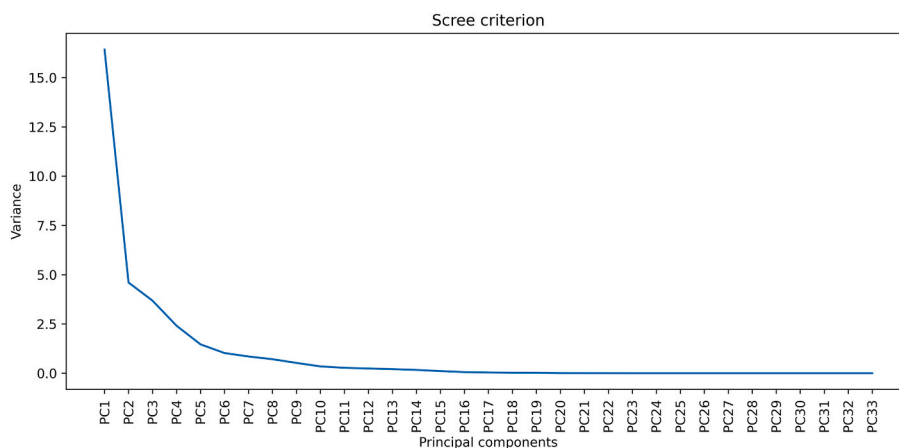


Fig. 6. Scree plot of PCA eigenvalues.

(blue) consists of 43 compounds, primarily composed of systems where Ca occupies the tetrahedral center, doped with Sr, and limited amounts of all other dopants (La, Bi), and mixture of Fe, Mn and Co in the octahedral center (Fig. 9c). Group 4 (green) consists of 30 compounds, primarily composed of systems where Ca occupies the tetrahedral center, doped with Ba, and lack of La-, Sr-dopants. The group 4 consist of mixture of Fe, Mn and Co in the octahedral center (Fig. 9d). Group 5 (grey) is the largest and most diverse cluster (76 samples - 31.7% of all compounds). It predominantly consists of brownmillerite-type structures doped with Ca and limited amounts of all other dopants (La, Bi, Sr) on the tetrahedral site and mixture of Fe, Mn and Co in the octahedral center (Fig. 9e).

3.6. Mechanistic interpretation of clusters and principal components

Principal Component 1 (PC1) differentiated clusters based on surface polarity and charge delocalization. Clusters 1, 2, 3, and 5 shared similar PC1 values and were characterized by.

- high value of (1) polar surface area, (2) molecular polarity index, (3) internal charge separation, (3,4) product of sigma and its overall variance, (5,6,7) electronic spatial extent of anion, cation and

neutral, (8) dipole moment, (9,10, 11) SOMO, LUMO and HOMO energy (Supplementary Materials, Table S2).

- low value of (1) nonpolar surface area, (2) balance of charge, (3) estimated density according to mass and volume, (4) volume (5,6,7) electronic energy of anion, cation and neutral, (8) electrophilicity; (9) electronegativity, (10) electron affinity and (11) ionization potential (Supplementary Materials, Table S2).

By contrast, Cluster 4 exhibited the inverse profile, with reduced polarity and increased structural compactness and stability indicators – suggesting lower catalytic activity but enhanced durability under SOEC operating conditions, characterized by.

- high value of (1) nonpolar surface area, (2) balance of charge, (3) estimated density according to mass and volume, (4) volume (5,6,7) electronic energy of anion, cation and neutral, (8) electrophilicity; (9) electronegativity, (10) electron affinity and (11) ionization potential (Supplementary Materials, Table S2).
- low value of (1) polar surface area, (2) molecular polarity index, (3) internal charge separation, (3,4) product of sigma and its overall variance, (5,6,7) electronic spatial extent of anion, cation and neutral, (8) dipole moment, (9,10, 11) SOMO, LUMO and HOMO energy (Supplementary Materials, Table S2).

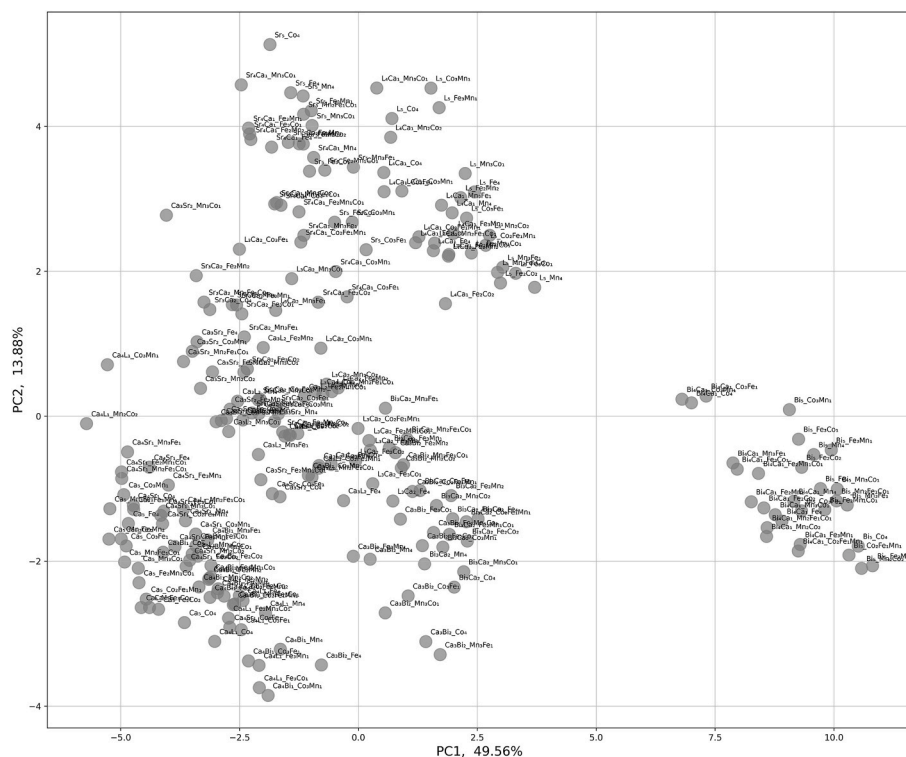


Fig. 7. PCA line map of the dataset onto the PC1-PC2 plane. Each point represents an individual compound.

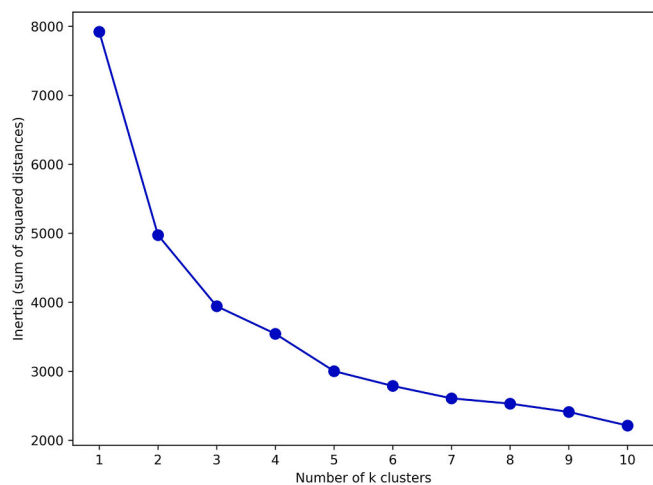


Fig. 8. Elbow plot for k-means clustering, showing the inertia (sum of squared distances) as a function of the number of clusters (k). The “elbow point” indicates the optimal number of clusters, where the decrease in inertia slows down significantly.

PC2 captured variations in the overall value of electrostatic potential ($\langle V_{\text{ESP}} \rangle$) across clusters (Fig. 10). Clusters 1, 4, and 5 showed high average values, while clusters 2 and 3 had higher values. This metric reflects the net electrostatic character of molecular surfaces, influencing water affinity and charge distribution in catalytic environments (Supplementary Materials, Table S2).

Overall, these findings indicate that clusters 1, 3, and 5 – composed predominantly of Ca-based, variably doped structures – exhibited the highest potential for catalytic activity based on electronic surface descriptors. Cluster 4, enriched in Bi and with minimal Ca, may serve as a thermally stable, oxidation-resistant candidate for prolonged SOEC use.

The *polar surface area* (PSA), Fig. 5 refers to the portion of a

molecule's solvent-accessible surface that is associated with polar fragments. It is computed by summing the van der Waals surface contributions of these fragments and serves as a proxy for the molecule's ability to engage in hydrogen bonding and dipole–dipole interactions. Originally used in medicinal chemistry to estimate membrane permeability, PSA has also found relevance in materials science for characterizing surface polarity and its effects on interfacial behavior. With respect to SOEC performance, PSA may provide insight into the material's surface affinity for polar species such as H_2O and OH^- , which are critical reactants and intermediates in steam electrolysis. A higher PSA typically implies increased hydrophilicity, potentially facilitating adsorption and dissociation of water molecules - key steps in the oxygen evolution reaction. However, an excessively polar surface may also promote undesired side reactions, such as surface hydroxylation or the formation of inactive surface species, especially under prolonged exposure to high-temperature steam. On a mechanistic level, the distribution and magnitude of PSA could modulate the balance between catalytic activity and structural stability. Optimizing PSA may therefore be crucial for tuning the anode's surface to simultaneously enable efficient electrochemical conversion and resist degradation processes. From this perspective, PSA is not merely a structural or electronic parameter but a functional descriptor reflecting the interplay between surface chemistry, gas reactivity, and material robustness in SOEC environments.

The *nonpolar surface area* (NPSA) that represents the portion of a molecule's solvent-accessible surface composed of electronically neutral and nonpolar fragments. It is calculated by excluding surface contributions from polar fragments or charged centers and thus provides insight into the spatial distribution of hydrophobic character on the molecular surface. NPSA is widely used in molecular science to evaluate hydrophobicity and surface-based interactions. From the perspective of SOEC operation, this descriptor may serve as an indirect indicator of surface chemical behavior, particularly in relation to the affinity for polar species such as H_2O or OH^- . A relatively high NPSA implies the presence of extended apolar domains, which could reduce susceptibility to surface hydroxylation and mitigate degradation under high-temperature steam electrolysis conditions. These hydrophobic regions may also facilitate

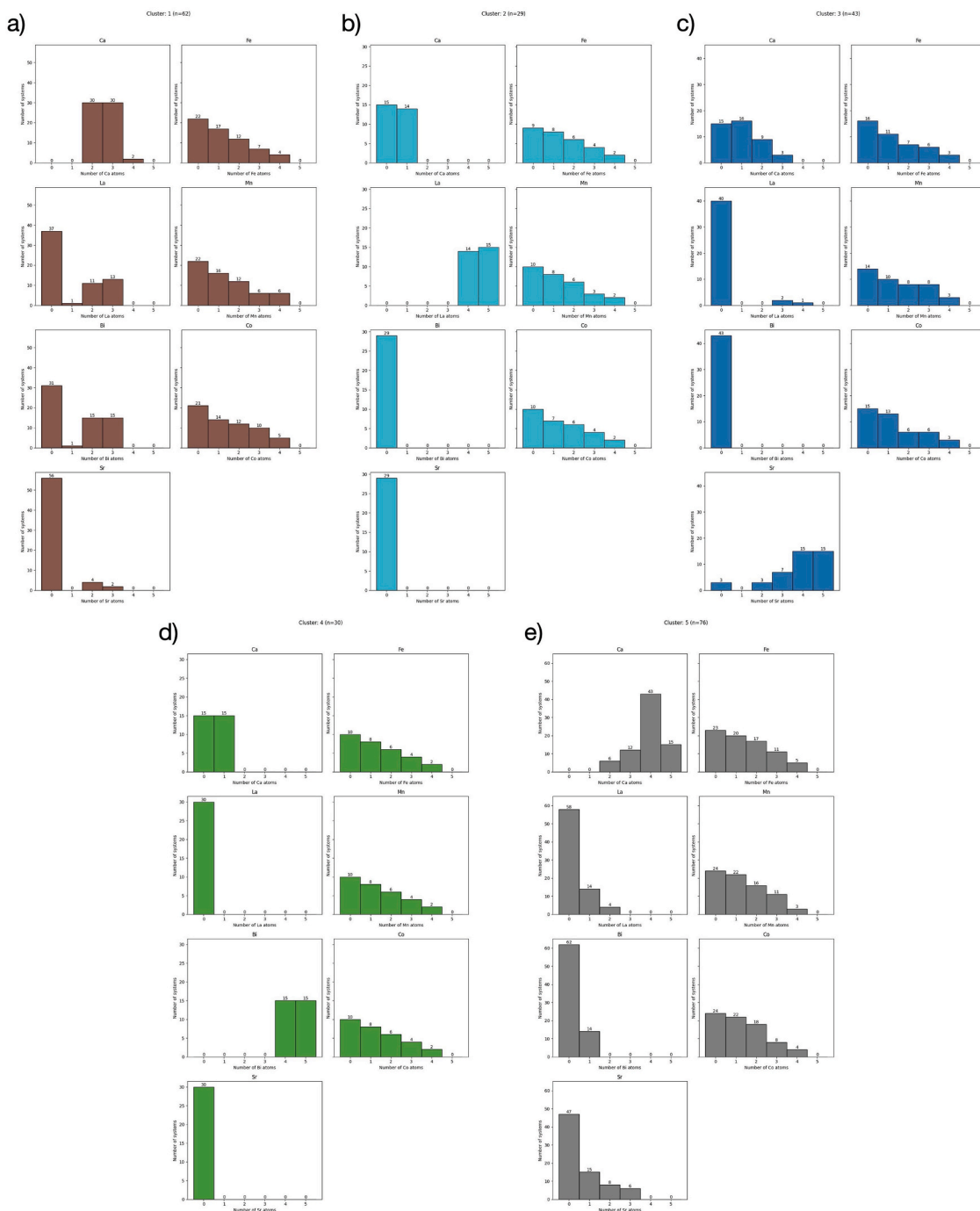


Fig. 9. Distribution of atomic compositions within clusters obtained from k-means analysis. Histograms represent the number of systems containing specific numbers of selected atoms within each cluster, illustrating compositional changes and variability among grouped compounds.

gas-phase transport by minimizing undesired water-surface interactions, which is particularly relevant for long-term operation in humid environments. From a mechanistic viewpoint, an elevated NPSA could reflect a favorable shift in the balance between polar and nonpolar interactions at the solid-gas interface, potentially influencing adsorption selectivity and catalytic behavior. Although this interpretation is partly speculative, it suggests that increased NPSA may support hydrogen evolution while reducing parasitic reactions, ultimately enhancing chemical durability and maintaining surface integrity during cyclic SOEC operation. In composite anodes, where both ionic and electronic conduction must be balanced, such surface characteristics may also contribute to improved interface stability and overall electrochemical

performance.

The *molecular polarity index* (MPI) is a surface-based descriptor that quantifies the overall polarity of a molecule by averaging the absolute value of the electrostatic potential (ESP) mapped onto the molecular van der Waals surface. Unlike the dipole moment, which captures only the net separation of charge, MPI reflects the full three-dimensional distribution of electrostatic potential, including contributions from localized polar regions. A higher MPI indicates a more pronounced presence of strong electrostatic features across the surface, corresponding to greater molecular polarity. Regarding the operation of SOEC anodes, MPI may provide a compact, yet informative measure of how the surface interacts with polar systems such as water and hydroxide ions - both of which are

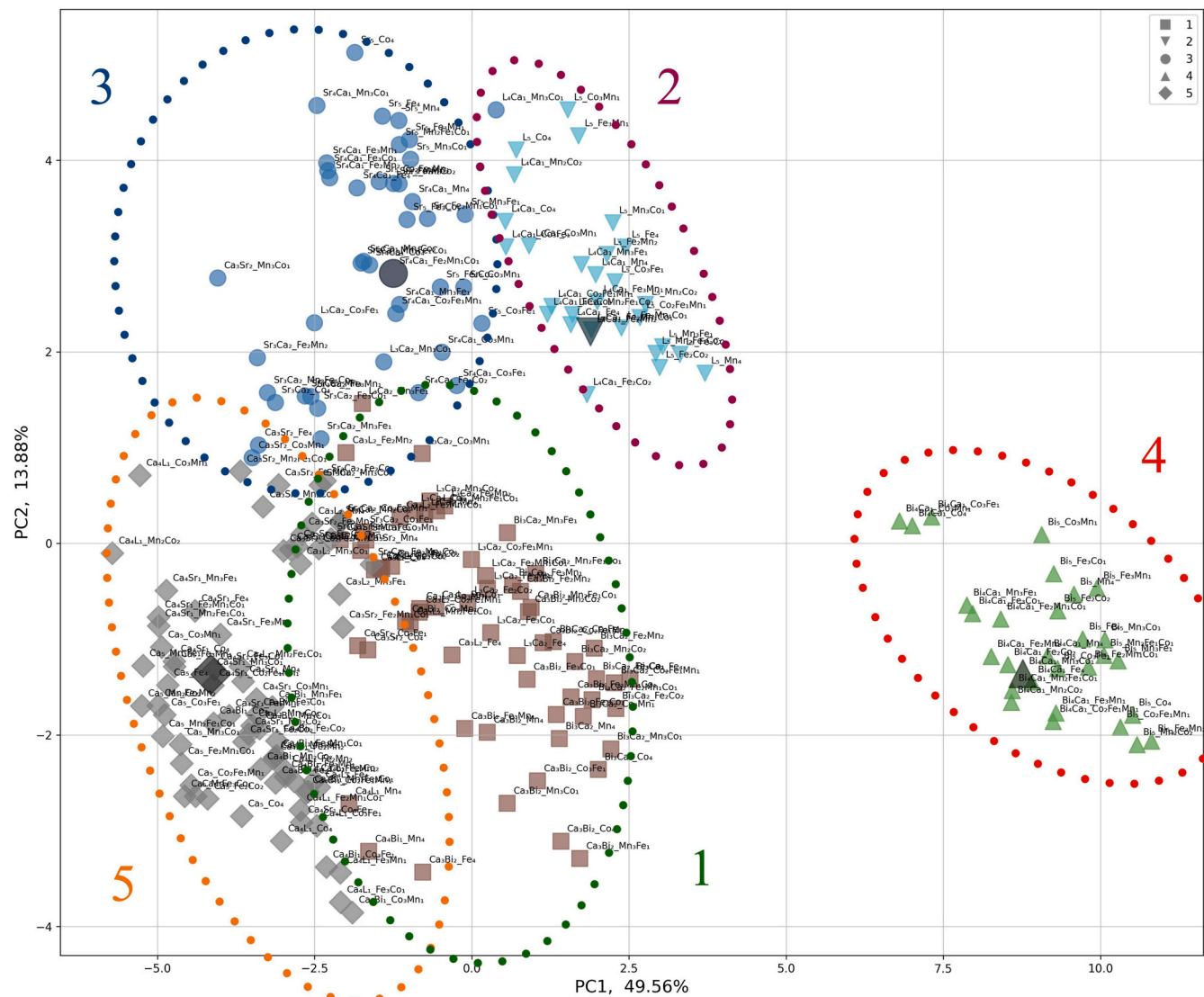


Fig. 10. PCA projection of the dataset onto the PC1-PC2 plane, with k-means clustering results overlaid. Each point represents an individual data instance, and colors and shapers indicate cluster assignments. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Distribution and main composition of compound clusters based on elemental doping patterns.

Cluster number	Number of compounds	Percent of total compounds	Main composition
1	62	25,83%	Ca- compounds doped with La and Bi (equal proportions)
2	29	12,08%	La- compounds doped with Ca
3	43	17,92%	Sr- compounds doped with Ca
4	30	12,50%	Bi- compounds doped with Ca
5	76	31,67%	Ca- compounds doped with La, Bi and Sr (Ca predominant)

central to the electrochemical processes occurring during steam electrolysis. Materials with elevated MPI values are likely to possess regions of strong local electric fields, which can enhance the adsorption and activation of polar species, thereby promoting catalytic steps such as water dissociation. However, such polarity may also correlate with increased vulnerability to irreversible surface transformations, including hydroxylation or hydration-induced structural degradation,

Table 2
Representative samples from each cluster.

Cluster number	Cluster representative sample
1	$\text{Ca}_3\text{La}_2\text{Co}_2\text{FeMnO}_{20}$
2	$\text{La}_4\text{CaFe}_2\text{Mn}_2\text{O}_{20}$
3	$\text{Sr}_4\text{CaFe}_2\text{MnCoO}_{20}$
4	$\text{Bi}_4\text{CaMn}_3\text{CoO}_{20}$
5	$\text{Ca}_4\text{SrMn}_3\text{CoO}_{20}$

especially under prolonged operation in steam-rich environments. In mechanistic terms, MPI offers a balance between reactivity and stability: moderate polarity can support efficient water splitting without overly enhancing parasitic surface reactions. Therefore, MPI emerges as a meaningful descriptor for screening candidate materials in terms of their surface electrostatics, reactivity toward steam, and potential resilience in high-temperature SOEC conditions. Water adsorption on oxide surfaces is strongly governed by local electrostatics and electrostatic potential-based descriptors have been shown to predict and rationalize adsorption energies of nucleophiles on oxide surfaces [38]. For perovskite oxides specifically, surface wetting is linked to surface chemistry and has been investigated with both experiment and DFT calculation.

This motivates the use of polarity/electrostatics descriptors in current study [39].

The *internal charge separation* (ICS), Fig. 5 quantifies the spatial displacement between regions of electron depletion (holes) and electron accumulation (electrons) within a molecule, typically computed as the distance between the centroids of positive and negative regions of the electron density difference between the excited and ground state. This descriptor reflects the degree of intramolecular charge redistribution upon excitation or perturbation. It is derived from the analysis of density difference grids and provides a direct measure of how far electrons shift within a system as a response to external stimuli such as electric fields or redox processes. When considering SOEC-relevant conditions, internal charge separation may be interpreted as an indicator of the material's ability to sustain or accommodate internal polarization under operational conditions. A larger value suggests a more pronounced shift of electron density within the structure, potentially enhancing the material's ability to separate charge carriers spatially - a property that may facilitate ionic and electronic transport at elevated temperatures. However, excessive internal charge separation could also correlate with structural strain or increased susceptibility to electronic instabilities, especially under long-term electrochemical cycling. From a mechanistic perspective, moderate internal charge separation may be beneficial for enhancing surface reactivity toward polar species such as H_2O by inducing localized electric fields, which in turn can lower the activation barriers for key surface processes such as dissociative adsorption or proton abstraction. In this light, internal charge separation serves not only as a geometric measure of electron redistribution but also as a functional descriptor connecting molecular electronic structure with macroscale electrochemical performance in high-temperature steam electrolysis.

The *overall variance* (σ_{total}^2) of the molecular electrostatic potential (Fig. 5) is a quantitative measure of how widely the surface ESP values deviate from their mean across the van der Waals surface. It is computed as the statistical variance of ESP at all surface grid points and reflects the extent to which the molecule exhibits localized regions of high positive and negative electrostatic potential. A larger σ_{total}^2 indicates greater heterogeneity in the ESP distribution and suggests the presence of distinct electrostatic domains capable of strong intermolecular interactions. From the perspective of SOEC operation, a high overall ESP variance may imply a surface that combines strongly positive and strongly negative regions, potentially enhancing adsorption selectivity toward polar species such as water or hydroxide. This spatial variation may facilitate surface polarization effects that promote key electrochemical steps, including water activation and proton-coupled electron transfer. However, such surfaces may also be prone to enhanced local reactivity, which - if not properly balanced - could increase susceptibility to surface degradation or parasitic reactions under harsh operational conditions. Grounded in mechanistic reasoning, σ_{total}^2 serves as a descriptor of the electrostatic 'texture' of the material surface, with potential implications for both catalytic performance and chemical resilience in steam electrolysis environments. Therefore, tuning the overall ESP variance may be key to designing anode materials with an optimal balance of reactivity and stability.

The *balance of charges* (ν), Fig. 5 is a scalar descriptor that quantifies the degree of symmetry in the distribution of electrostatic potential across the molecular van der Waals surface. It is defined as a normalized index comparing the magnitudes and spatial extents of regions with positive and negative electrostatic potential. A low ν value indicates a well-balanced distribution of opposing electrostatic regions, while a high ν suggests a skewed or asymmetric ESP landscape dominated by one polarity. In the context of SOEC functionality, the balance of charges may reflect how evenly the surface can engage in both electrophilic and nucleophilic interactions, which is critical for reversible processes such as steam splitting and hydrogen evolution. A highly unbalanced ESP surface may favor only one type of interaction, potentially enhancing

selectivity but reducing versatility. Conversely, a well-balanced surface (low ν) could support dual functionality - adsorbing both electron-rich and electron-deficient species - which is advantageous under dynamic SOEC conditions involving redox cycling. In light of molecular-level mechanisms, an optimal charge balance may contribute to surface stability by preventing localized charge accumulation that could trigger parasitic reactions or structural degradation. It may also facilitate smoother transitions between anodic and cathodic regimes, especially in reversible operation modes. Thus, the ν index offers a compact yet informative descriptor of electrostatic symmetry, linking molecular surface properties to the broader electrochemical robustness and adaptability of SOEC anode materials.

The *product of σ_{total}^2 and ν* ($\sigma_{\text{total}}^2 \cdot \nu$), Fig. 5 represents a composite descriptor that integrates the variability and symmetry of the molecular electrostatic potential over the van der Waals surface. Here, σ_{total}^2 quantifies the overall variance in ESP, reflecting the magnitude of electrostatic heterogeneity, while ν measures the degree of balance between positive and negative ESP regions. Their product captures both how pronounced and how symmetrically distributed the electrostatic features are across the molecular surface. With respect to SOEC performance, a high value of $\sigma_{\text{total}}^2 \cdot \nu$ may indicate a surface with well-developed and balanced electrostatic domains, capable of engaging in strong yet non-preferential interactions with both electron-donating and electron-withdrawing species. This duality could be particularly advantageous under dynamic operational conditions, where the surface must facilitate both oxidation and reduction processes, or interact with intermediates of varying polarity. From a mechanistic standpoint, such a surface may support efficient adsorption of polar reactants like H_2O while simultaneously minimizing localization of charge that could otherwise lead to surface degradation or passivation. Thus, this descriptor provides a holistic measure of the surface's electrostatic 'fitness' - combining energetic intensity with spatial neutrality - and may serve as a valuable criterion for identifying anode materials with balanced reactivity and robust long-term electrochemical performance in steam electrolysis applications.

The *molecular volume* (V), Fig. 5 is a spatial descriptor representing the three-dimensional size of a molecule, typically defined as the volume enclosed by the 0.001 a.u. electron density isosurface. It captures both the geometrical extension and the compactness of the molecular framework. On the other hand, the *estimated density* is a descriptor derived by dividing the molecular mass by this calculated volume, which approximates the density the material would exhibit in the condensed phase and serves as a proxy for packing efficiency and mass distribution. Regarding the operation of SOEC anodes, these two descriptors may reflect structural traits with practical implications for mechanical integrity, thermal expansion, and electrochemical compatibility. Materials with lower molecular volume and correspondingly higher estimated density are often associated with tighter atomic packing, which may enhance sintering behavior, improve interfacial contact with other cell components, and support stability under high-temperature conditions. Conversely, more voluminous structures may allow for greater structural flexibility or facilitate diffusion pathways for gas-phase species such as H_2 or H_2O . On a mechanistic level, the balance between molecular volume and density may influence how dopants are accommodated within the lattice, how defects propagate under electrochemical cycling, and how the material tolerates redox-driven volume changes. A favorable density profile might also reduce porosity or phase segregation in composite anodes. Thus, while volume and density are fundamentally geometric and gravimetric parameters, their integration offers valuable insight into the physical robustness and operational viability of SOEC anode materials.

The *electronic spatial extent* ($\langle r^2 \rangle$), Fig. 5 is a quantum mechanical descriptor that quantifies the average square distance of the electron density from the molecular centroid. It provides a direct measure of the spatial 'size' or delocalization of the electronic 'cloud' and is particularly

sensitive to the overall distribution of valence electrons. When considering SOEC-relevant conditions, the $\langle r^2 \rangle$ value computed for the neutral species offers insight into how compact or diffuse the electron density is under ground-state conditions. A lower electronic spatial extent typically implies a more localized and tightly bound electron distribution, which may correlate with enhanced chemical stability and reduced susceptibility to electronic disorder or polarizability under operating conditions. Conversely, a higher $\langle r^2 \rangle$ in the neutral system may reflect a more delocalized electron density, potentially facilitating long-range electronic transport or improving interaction with extended lattice environments - features that may be beneficial for anode performance under high-temperature conditions. With mechanistic relevance, moderate electronic spatial extent may strike a balance between reactivity and structural coherence, supporting efficient charge transfer without compromising lattice stability. In addition to the neutral state, the spatial extents of the corresponding cationic and anionic species were also evaluated. These provide auxiliary insight into the system's response to electron removal or addition. A substantial increase in $\langle r^2 \rangle$ upon electron addition (anion) may suggest diffuse LUMO character and limited electron accommodation capability, whereas a large contraction upon ionization (cation) may indicate HOMO localization. While these charged states are not directly realized in the functioning SOEC anode, their comparison with the neutral state can illuminate the system's charge redistribution tendencies and potential vulnerability to oxidative or reductive extremes under dynamic electrochemical cycling.

The *dipole and quadrupole moments* (Fig. 5) are classical multipole descriptors characterizing the deviation of electronic charge from spherical symmetry. The dipole moment quantifies the net separation of charge, while the quadrupole moment captures more complex charge anisotropies. Together, they help characterize the electrostatic landscape of a molecule and provide insights into how SOEC anode surfaces may engage with charged species and external fields. Their combined effect may influence adsorption, polarization, and the electrochemical performance of the material. The quadrupole moment, on the other hand, offers a more nuanced picture of how charge is distributed over molecular planes or elongated structural motifs, which is particularly relevant in layered or anisotropic oxide frameworks. It may correlate with the ability of the material to stabilize electric field gradients, influencing properties such as lattice distortion, defect accommodation, or interaction with the electrolyte. Taken together, these multipolar descriptors help characterize not just the magnitude but also the spatial character of molecular electrostatics. Their combined analysis may therefore provide insight into how anode materials engage with charged species, respond to external fields, and contribute to overall electrochemical performance in SOEC applications.

The *energies of the HOMO and LUMO orbitals* (Fig. 5) define the frontier electronic structure and are essential for understanding the redox behavior of a material. The HOMO energy indicates susceptibility to oxidation, while the LUMO reflects the ability to accept electrons. Their relative positions also determine the HOMO–LUMO gap, often correlated with electronic softness. In SOEC anodes, these energies influence charge transfer processes, surface reactivity, and the balance between catalytic performance and structural stability.

Absolute electronegativity (Fig. 5) also known as Mulliken electronegativity, is a global quantum chemical descriptor that reflects the overall tendency of a system to attract and retain electron density. It is derived from fundamental electronic properties and provides a measure of the average energy required to either donate or accept an electron. As such, it serves as a useful indicator of the electronic character and redox behavior of a material. From the perspective of SOEC operation, absolute electronegativity can offer insight into how the material surface may interact with charged species, particularly those involved in steam electrolysis, such as hydroxide ions or water molecules. A material with higher electronegativity is generally more likely to attract electrons, potentially enhancing the adsorption of electron-rich intermediates or

facilitating surface polarization. This may support efficient catalytic processes, such as water dissociation, which require localized electron density shifts. However, high electronegativity could also indicate a susceptibility to over-accumulation of charge or reduced stability under strongly oxidative conditions, depending on the specific lattice and bonding environment. Materials with lower electronegativity may, by contrast, show improved resilience against such effects but could be less effective in stabilizing key reaction intermediates. Thus, absolute electronegativity provides a compact, meaningful measure of how the electronic nature of the anode influences both its catalytic performance and its long-term electrochemical stability in high-temperature SOEC operation.

The *ionization potential* (IP), Fig. 5 is a fundamental quantum chemical descriptor that reflects the energy required to remove an electron from a neutral system, thereby generating a cationic form. It serves as a direct measure of how tightly the highest-energy electrons are bound and is closely related to the position of the HOMO level. A high ionization potential implies that the molecule is electronically stable and resistant to oxidation, while a low IP suggests that electron removal is energetically more favorable. In the context of SOEC functionality, the ionization potential provides insight into the material's susceptibility to oxidative processes under anodic conditions. Anodes with lower IP values may more readily donate electrons, potentially enhancing catalytic activity during the oxidation of water and the evolution of oxygen. However, such behavior might come at the cost of decreased chemical robustness, especially under sustained high-temperature operation where surface degradation is a concern. Conversely, materials with higher ionization potential tend to be more resistant to electron loss, which can translate to improved oxidative stability and longer operational lifetimes. Grounded in mechanistic reasoning, the ionization potential helps predict how readily the anode material can participate in charge-transfer processes at the electrode–electrolyte interface, and how it might balance between reactivity and structural durability in the harsh conditions typical of SOEC environments.

The *electron affinity* (EA), Fig. 5 is a quantum chemical descriptor that represents the energy change associated with the addition of an electron to a neutral system, forming an anion. It reflects the ability of a material to accommodate extra electron density and is directly linked to the position and nature of the LUMO. A higher electron affinity indicates a greater tendency to accept electrons, often associated with increased surface reactivity toward electron-rich species or with the stabilization of charge-separated states. With respect to SOEC performance, electron affinity offers insight into how the surface might interact with negatively charged intermediates, such as hydroxide ions, or how it may respond under transient reductive conditions, for instance during SOEC operation. A moderate electron affinity may support catalytic activity by enabling temporary electron acceptance, which can facilitate crucial steps like water adsorption and activation. However, an overly high EA might also raise concerns about unwanted charge trapping or structural strain due to localized reduction. Conversely, low electron affinity suggests a limited ability to stabilize excess electron density, which could protect the material from over-reduction but might also hinder its involvement in certain catalytic pathways. From a mechanistic perspective, electron affinity provides a useful indicator of how the anode surface balances reactivity and electronic resilience - key factors that influence both efficiency and durability in high-temperature electrochemical environments typical of SOEC systems.

Among with PC2 the *overall averaged value of electrostatic potential* ($\langle V_{ESP} \rangle$) is a scalar descriptor representing the mean electrostatic potential calculated over the molecular van der Waals surface. Unlike the variance or polarity indices, which emphasize local fluctuations or asymmetries, $\langle V_{ESP} \rangle$ captures the net electrostatic character of the entire surface and reflects the general tendency of the molecule to present an attractive or repulsive field to surrounding species. A positive $\langle V_{ESP} \rangle$ implies a globally electron-deficient surface, potentially favoring

interactions with electron-rich species, while a negative $\langle V_{\text{ESP}} \rangle$ indicates a more electron-rich environment. With regard to SOEC anode materials, the overall averaged ESP may offer a compact metric for anticipating how the surface engages with polarizable or charged intermediates, such as water molecules or hydroxide ions. For instance, a surface with moderately negative $\langle V_{\text{ESP}} \rangle$ values may enhance the physisorption or activation of H_2O through electrostatic attraction, thereby supporting efficient catalytic turnover in the steam-splitting process. Conversely, surfaces with strongly positive $\langle V_{\text{ESP}} \rangle$ might exhibit diminished affinity toward such polar species but improved resistance to over-reduction or charge accumulation. From a mechanistic standpoint, $\langle V_{\text{ESP}} \rangle$ serves as a baseline indicator of the surface's net electrostatic disposition, potentially modulating long-range interactions prior to specific chemical adsorption events. While less spatially resolved than other descriptors, it complements more localized metrics like MPI or σ_{total}^2 by offering an integrated perspective on surface reactivity. In this context, tuning $\langle V_{\text{ESP}} \rangle$ may help strike a balance between catalytic responsiveness and resistance to undesirable surface polarization or degradation under SOEC operating conditions.

3.7. Selection of representatives of clusters

To identify structurally and electronically representative candidates for further analysis, we applied a hybrid k-means with medoid approach to each of the five clusters derived from PCA-reduced descriptor space. Medoid selection ensures that each chosen compound is not an abstract centroid but a real, chemically plausible structure within the dataset – sharing maximal similarity to its cluster while preserving structural uniqueness.

The selected representatives are marked as enlarged, darkened points in the PCA projection (Fig. 8). Each medoid serves as a prototype for its respective cluster, encapsulating key trends in dopant distribution and descriptor profiles. Table 2 summarizes these representatives.

3.7.1. Cluster-specific insights

- Clusters 1, 3, and 5: These groups are dominated by Ca-based systems doped at the A-site with varying proportions of La, Bi, and Sr. They consistently exhibit elevated polar surface area (PSA), molecular polarity index (MPI), internal charge separation (ICS), and electronic spatial extent, all indicative of high surface reactivity and enhanced catalytic potential. These compounds emerge as prime candidates for SOEC anodes focused on water activation and rapid redox cycling.
- Cluster 2: Characterized by intermediate descriptor values, limited to La and Ca doping with minimal Bi or Sr presence. It likely represents a transitional regime between high-polarity, reactive systems and more electronically stable variants. This group may be suitable where moderate catalytic activity is balanced with structural robustness.
- Cluster 4: Comprised primarily of Bi-rich, Ca-poor compositions, this cluster demonstrates low PSA and MPI, high ESP imbalance, and limited charge delocalization. While less active catalytically, such compounds may offer superior thermal and oxidative stability, positioning them as durable anode candidates under harsh SOEC conditions.

This cluster-based virtual screening thus enables the rational selection of optimal materials for further theoretical or experimental validation. By leveraging medoid-based sampling, we ensure that each representative embodies the essential physicochemical traits of its group – allowing researchers to prioritize candidates with tailored combinations of reactivity, polarity, and long-term stability. From a sustainability and SSbD perspective, the same transparent clustering-and-medoid strategy acts as an early-tier sustainability filter, because it reduces the candidate set to a small, chemically diverse and auditable

subset for subsequent, more comprehensive sustainability assessment (e.g., criticality, safety, and environmental footprint) while preserving the rationale behind prioritization.

For complete clustering statistics and atomic composition distributions, refer to [Supplementary Table S3](#).

4. Conclusion

In this work, the descriptor based analysis of fragment level DFT data is intended as a first pass, high throughput screening within a single brownmillerite family, not as a quantitative prediction of SOEC anode performance. The mechanistic interpretation offered throughout the manuscript should therefore be viewed as literature grounded, hypothesis generating reasoning that links dopant induced changes in local metal oxygen bonding and redox flexibility to processes relevant for oxygen electrodes. Key SOEC level determinants, including oxygen vacancy formation energies and equilibrium non-stoichiometry, vacancy migration barriers (ionic transport), and extended electronic structure (band dispersion and long range carrier transport), are outside the scope of the present fragment representation and must be assessed using periodic approaches and, ultimately, experimental measurements. Accordingly, the PCA and clustering results are used to prioritize chemically diverse candidate compositions for subsequent validation, rather than to claim proven mechanisms or validated performance rankings under operating SOEC conditions.

Representative compounds identified via medoid selection from each cluster offer promising pathways for experimental validation. Overall, this work establishes a scalable, AI-augmented computational approach to rational anode design, providing a robust framework for accelerating material discovery and supporting the development of efficient, durable, and sustainable SOEC technologies.

Moreover, when these findings are placed in the context of existing literature, it becomes evident that some compounds with similar qualitative and quantitative compositions have already been investigated. By contrast, several of the compounds proposed here currently exist only as theoretical candidates, and we therefore encourage experimental groups to attempt their synthesis, provided that no methodological constraints preclude it.

The candidate proposed from the analysis of cluster 1 is the compound $\text{Ca}_3\text{La}_2\text{Co}_2\text{FeMnO}_{20}$ (LCCFM), which is composed of calcium, lanthanum, cobalt, iron, and manganese. A related composition $\text{La}_{0.6}\text{Ca}_{0.4}\text{Co}_{0.2}\text{Fe}_{0.8}\text{O}_3$ (LCCF), featuring inverted A-site doping ratios and a non-equimolar distribution of B-site cations, has recently been investigated by Rashid et al. [40]. Since the compound identified in our study currently exists only as a theoretical candidate, we recommend that experimental efforts be directed toward the attempted synthesis of the corresponding triple brownmillerite phase, particularly in view of literature reports indicating that such structures are attainable [41,42].

The second candidate is the compound $\text{La}_4\text{CaFe}_2\text{Mn}_2\text{O}_{20}$. A closely related phase has already been characterized in recent years [43,44]. However, in the previously studied materials the dopant ratios at both the A-site and the B-site were equimolar, whereas in the proposed composition the A-site sublattice is enriched in lanthanum. We therefore put forward this oxide as a candidate material in which moderate electrochemical activity is compensated by high structural robustness.

Within cluster 3, the proposed composition, $\text{Sr}_4\text{CaFe}_2\text{MnCoO}_{20}$, is a La free, Sr based perovskite in which calcium is introduced at the A-site, while the B-site contains iron co doped with manganese and cobalt in equal proportions. This material is entirely lanthanum free and relies exclusively on readily available elements, which is particularly advantageous in the context of the European Union's raw materials strategy. A closely related oxide, featuring A-site substitution only, has been comprehensively characterized by Ede et al. [45].

Within cluster 4, the proposed representative composition is $\text{Bi}_4\text{CaMn}_3\text{CoO}_{20}$ a bismuth and manganese-based oxide featuring calcium substitution at the A-site and cobalt substitution at the B-site. This

compound is entirely lanthanum free. Related structures were synthesized in 2001 by Qin et al. [46], although in those systems the doping was confined to the A-site. We therefore suggest this compound as a promising candidate for a structurally durable anode material.

The final composition proposed for experimental investigation is $\text{Ca}_4\text{SrMn}_3\text{CoO}_{20}$. This oxide is likewise La free and exhibits dual A-site doping with small amounts of strontium, together with a minor cobalt substitution at the B-site. A closely related phase was synthesized by Boulahya et al. [47] as early as 2007, although in that case the A-site cation ratios were inverted relative to the composition proposed here.

From the perspective of the Safe and Sustainable by Design framework, the ideal materials are non-toxic, readily recyclable compounds composed exclusively of elements not included on the Critical Raw Materials list [48]. In the context of SOEC anodes, pure CFO (brownmillerite $\text{Ca}_2\text{Fe}_2\text{O}_5$), formally meets these criteria; however, its electrochemical properties are only moderate [49,50], which necessitates performance enhancement through aliovalent doping or composite formation [51–53] if it is to be employed outside the narrow window of intermediate temperature SOEC (IT SOEC) operation. This, in turn, motivates the consideration of compositions that incorporate elements classified as Critical Raw Materials under current EU definitions. Even in this space, the design flexibility remains substantial, as the availability and environmental burden of different elements vary significantly [54–57]. It is therefore reasonable to allow the use of elements whose environmental impact, cost, and criticality are comparatively low, specifically La, Sr, and Bi.

This proof-of-concept workflow is intended as an early-stage comparative screening within a fixed DFT protocol and does not capture bulk/transport properties or constitute a comprehensive SSbD assessment, which requires additional data beyond the scope of this work.

CRedit authorship contribution statement

Dawid Falkowski: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Alicja Mikolajczyk:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Data curation, Conceptualization. **Piotr Skurski:** Writing – review & editing, Writing – original draft, Software, Investigation. **Jakub Brzeski:** Software, Investigation. **Tomasz Puzyn:** Writing – review & editing, Supervision, Project administration.

Declaration of competing 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.

Acknowledgments

This research was supported by the Polish Ministry of Science and Higher Education grant Doktorat Wdrożeniowy 2023, grant agreement no. DWD/7/0321/2023; European Union's Horizon Europe Digital, Industry and Space programme, within NOUVEAU project, grant no. 10105878, and University of Gdansk UGrants-start (no. 1200.6010.84.2025). All quantum-chemical calculations have been carried out using resources provided by Wrocław Centre for Networking and Supercomputing (<https://wcss.pl>) grants no. 455 (P. S.) and 560 (J. B.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijhydene.2026.154543>.

Data availability

All data supporting the findings of this study, including descriptor values and clustering results, are provided in the Supplementary Materials. Additional raw data are available from the corresponding author upon reasonable request.

References

- [1] European Parliament. Key enabling technologies for Europe's technological sovereignty. <https://doi.org/10.2861/24482>; 2021.
- [2] European Commission. The European green deal. 2019.
- [3] European Commission. REPowerEU plan. 2022.
- [4] European Parliament. Revision of the energy performance of buildings directive. 2024.
- [5] Risco-Bravo A, Varela C, Bartels J, Zondervan E. From green hydrogen to electricity: a review on recent advances, challenges, and opportunities on power-to-hydrogen-to-power systems. *Renew Sustain Energy Rev* 2024;189:113930. <https://doi.org/10.1016/j.rser.2023.113930>.
- [6] European Commission. Advanced materials for industrial leadership. 2024.
- [7] Wolf SE, Winterhalder FE, Vibhu V, de Haart LGJ, Bert, Guillon O, Eichel R-A, et al. Solid oxide electrolysis cells – current material development and industrial application. *J Mater Chem A Mater* 2023;11:17977–8028. <https://doi.org/10.1039/D3TA02161K>.
- [8] Wang Y, Li W, Ma L, Li W, Liu X. Degradation of solid oxide electrolysis cells: phenomena, mechanisms, and emerging mitigation strategies—A review. *J Mater Sci Technol* 2020;55:35–55. <https://doi.org/10.1016/j.jmst.2019.07.026>.
- [9] Sumi H, Shimada H, Yamaguchi Y, Nomura K, Sato K. Why is the performance different between small- and large-scale SOFCs? *Electrochim Acta* 2023;443:141965. <https://doi.org/10.1016/j.electacta.2023.141965>.
- [10] Gandiglio M, Marocco P, Nieminen A, Santarelli M, Kiviaho J. Energy and environmental performance from field operation of commercial-scale SOFC systems. *Int J Hydrogen Energy* 2024;85:997–1009. <https://doi.org/10.1016/J.IJHYDENE.2024.08.332>.
- [11] Cai Q, Luna-Ortiz E, Adjiman CS, Brandon NP. The effects of operating conditions on the performance of a solid oxide steam electrolyser: a model-based study. *Fuel Cells* 2010;10:1114–28. <https://doi.org/10.1002/fuce.200900211>.
- [12] Gaikwad PS, Mondal K, Shin YK, van Duin ACT, Pawar G. Enhancing the Faradaic efficiency of solid oxide electrolysis cells: progress and perspective. *npj Comput Mater* 2023;9:149. <https://doi.org/10.1038/s41524-023-01044-1>.
- [13] Kamlungsua K, Su PC, Chan SH. Hydrogen generation using solid oxide electrolysis cells. *Fuel Cells* 2020;20:644–9. <https://doi.org/10.1002/FUCE.202070602>.
- [14] Bedon A, Viricelle JP, Rieu M, Mascotto S, Glisenti A. Single chamber solid oxide fuel cells selective electrodes: a real chance with brownmillerite-based nanocomposites. *Int J Hydrogen Energy* 2021;46:14735–47. <https://doi.org/10.1016/J.IJHYDENE.2021.01.220>.
- [15] Hauch A, Blennow PG, Dalby KN, Drasbæk DB, Heireld-Clausen T, Padinjarethil AK, et al. The topsoe perspective: from electrode nanostructures to MW scaled SOEC systems. *ECS Trans* 2023;111:1125. <https://doi.org/10.1149/11106.1125ECST>.
- [16] Zong S, Zhao X, Jewell LL, Zhang Y, Liu X. Advances and challenges with SOEC high temperature co-electrolysis of $\text{CO}_2/\text{H}_2\text{O}$: materials development and technological design. *Carbon Capture Sci Technol* 2024;12:100234. <https://doi.org/10.1016/j.cst.2024.100234>.
- [17] Materials 2030 Initiative Contributors. Materials 2030 roadmap. 2022.
- [18] Mishra AK, Rajput S, Karamta M, Mukhopadhyay I. Exploring the possibility of machine learning for predicting ionic conductivity of solid-state electrolytes. *ACS Omega* 2023;8:16419–27. <https://doi.org/10.1021/ACSOMEGA.3C01400>.
- [19] Zhang Y, Xu X. Modeling oxygen ionic conductivities of ABO₃ perovskites through machine learning. *Chem Phys* 2022;558:111511. <https://doi.org/10.1016/j.chemphys.2022.111511>.
- [20] Behara S, Poonawala T, Thomas T. Crystal structure classification in ABO₃ perovskites via machine learning. *Comput Mater Sci* 2021;188:110191. <https://doi.org/10.1016/J.COMMATSCI.2020.110191>.
- [21] Gong C, Liu J, Dai S, Hao H, Liu H. Machine learning assisted prediction of the phonon cutoff frequency of ABO₃ perovskite materials. *Comput Mater Sci* 2024;239:112943. <https://doi.org/10.1016/J.COMMATSCI.2024.112943>.
- [22] Ahmad S, Emmery D, Gallucci F, van der Schaaf J. Neural network based molecular dynamics simulations for oxide ion transport in solid oxide electrolysis cell materials. *Int J Hydrogen Energy* 2025;180:151714. <https://doi.org/10.1016/J.IJHYDENE.2025.151714>.
- [23] Dieb TM, Ju S, Shiomi J, Tsuda K. Monte Carlo tree search for materials design and discovery n.d. <https://doi.org/10.1557/mrc.2019.40>.
- [24] Singh V, Patra S, Murugan NA, Toncu DC, Tiwari A. Recent trends in computational tools and data-driven modeling for advanced materials. *Mater Adv* 2022;3:4069–87. <https://doi.org/10.1039/D2MA00067A>.
- [25] Bayat M, Dong W, Thorborg J, To AC, Hattel JH. A review of multi-scale and multi-physics simulations of metal additive manufacturing processes with focus on modeling strategies. *Addit Manuf* 2021;47:102278. <https://doi.org/10.1016/J.ADDMA.2021.102278>.
- [26] Commission E, Centre JR, Caldeira C, Farcas L, Garmendia Aguirre I, Mancini L, et al. Safe and sustainable by design chemicals and materials – framework for the definition of criteria and evaluation procedure for chemicals and materials.

- Publications Office of the European Union; 2022. <https://doi.org/10.2760/487955>.
- [27] European Commission. Commission recommendation (EU) 2022/2510 of 8 December 2022 establishing a European assessment framework for 'safe and sustainable by design' chemicals and materials. 2022.
- [28] Garmendia Aguirre I, Abbate E, Bracalente G, Mancini L, Cappucci GM, Tosches D, et al. Safe and sustainable by design chemicals and materials. Revised framework (2025); 2025. <https://doi.org/10.2760/5103>.
- [29] Zhao Z, Wang X, Tang S, Cheng M, Shao Z. High-performance oxygen electrode Ce_{0.9}Co_{0.1}O_{2-δ}-LSM-YSZ for hydrogen production by solid oxide electrolysis cells. *Int J Hydrogen Energy* 2021;46:25332–40. <https://doi.org/10.1016/j.ijhydene.2021.05.072>.
- [30] Yu HS, He X, Li SL, Truhlar DG. MN15: a Kohn–Sham global-hybrid exchange–correlation density functional with broad accuracy for multi-reference and single-reference systems and noncovalent interactions. *Chem Sci* 2016;7: 5032–51. <https://doi.org/10.1039/C6SC00705H>.
- [31] Hay PJ, Wadt WR. *Ab initio* effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *J Chem Phys* 1985;82:270–83. <https://doi.org/10.1063/1.448799>.
- [32] Ditchfield R, Hehre WJ, Pople JA. Self-consistent molecular-orbital methods. IX. An extended gaussian-type basis for molecular-orbital studies of organic molecules. *J Chem Phys* 1971;54:724–8. <https://doi.org/10.1063/1.1674902>.
- [33] Gaussian R, Trucks G, Schlegel H, Scuseria G, Robb M, Cheeseman J, et al. *Gaussian*. Wallingford CT: Gaussian, Inc.; 2004. Wallingford, CT: Gaussian, Inc.
- [34] Lu T. A comprehensive electron wavefunction analysis toolbox for chemists, multiwfn. *J Chem Phys* 2024;161. <https://doi.org/10.1063/5.0216272/3309709>.
- [35] Pearson KLIII. On lines and planes of closest fit to systems of points in space. *London, Edinburgh Dublin Phil Mag J Sci* 1901;2:559–72. <https://doi.org/10.1080/14786440109462720>.
- [36] MacQueen J. Some methods for classification and analysis of multivariate observations. 1967.
- [37] Kaufmann L, Rousseeuw P. Clustering by means of medoids. *Data analysis based on the L1-Norm and related methods*. 1987. p. 405–16.
- [38] Halldin Stenlid J, Johansson AJ, Brinck T. The local electron attachment energy and the electrostatic potential as descriptors of surface–adsorbate interactions. *Phys Chem Chem Phys* 2019;21:17001–9. <https://doi.org/10.1039/C9CP03099A>.
- [39] Stoerzinger KA, Hong WT, Azimi G, Giordano L, Lee Y-L, Crumlin EJ, et al. Reactivity of perovskites with water: role of hydroxylation in wetting and implications for oxygen electrocatalysis. *J Phys Chem C* 2015;119:18504–12. <https://doi.org/10.1021/acs.jpcc.5b06621>.
- [40] Rashid A, Thiem M, Wellmann M, Bresser M, Lindemann O, Wiegers K-S, et al. Enhancing the hydrogen tolerance of La_{0.6}Ca_{0.4}Co_{0.2}Fe_{0.8}O_{3-δ} oxygen transport membranes with the substitution of 10 % Mn at the B site for plasma assisted CO₂ conversion. *Open Ceramics* 2025;22:100781. <https://doi.org/10.1016/j.oceram.2025.100781>.
- [41] Malveiro J, Ramos T, Ferreira LP, Waerenborgh JC, Nunes MR, Godinho M, et al. Magnesium doping on brownmillerite Ca₂FeAlO₅. *J Solid State Chem* 2007;180: 1863–74. <https://doi.org/10.1016/j.jssc.2007.04.009>.
- [42] Carvalho MD, Ferreira LP, Waerenborgh JC, Tsipis E, Lopes AB, Godinho M. Structure and magnetic properties of Ca₂Fe_{1-x}MnxAlO_{5+δ}. *J Solid State Chem* 2008;181:2530–41. <https://doi.org/10.1016/j.jssc.2008.06.008>.
- [43] Gerasimov EY, Rogov VA, Prosvirin IP, Isupova LA, Tsybulya SV. Microstructural changes in La_{0.5}Ca_{0.5}Mn_{0.5}Fe_{0.5}O₃ solid solutions under the influence of catalytic reaction of methane combustion. *Catalysts* 2019;9:563. <https://doi.org/10.3390/CATAL9060563>. 2019;9:563.
- [44] Rueda DA, Ibáñez-Sotelo JS, Parra-Mesa LV, Hernández-Valero AF, Sarmiento Vanegas JA, Deluque Toro CE, et al. Evidence of multifunctional properties of ferromanganite La_{0.5}Ca_{0.5}Mn_{0.5}Fe_{0.5}O₃. *J Mol Struct* 2024;1314:138821. <https://doi.org/10.1016/J.MOLSTRUC.2024.138821>.
- [45] Ede SR, Collins CN, Posada CD, George G, Wu H, Ratcliff WD, et al. Intermediate Sr₂Co_{1.5}Fe_{0.5}O_{6-δ} tetragonal structure between perovskite and brownmillerite as a model catalyst with layered oxygen deficiency for enhanced electrochemical water oxidation. *ACS Catal* 2021;11:4327–37. <https://doi.org/10.1021/acscatal.1c00465>.
- [46] Qin YL, García-Muñoz JL, Zandbergen HW, Casais MT. A-face-centered Bi_{0.75}Ca_{3.25}Mn₃O₁₀: a Bi-based orthorhombic 4310-type bilayered manganite. *Phys Rev B* 2001;64:172405. <https://doi.org/10.1103/PhysRevB.64.172405>.
- [47] Boulahya K, Hernando M, Parras M, González-Calbet JM. New stabilized phases in the Sr/Ca–Mn–Co–O system: structural–magnetic properties relationship. *J Mater Chem* 2007;17:1620–6. <https://doi.org/10.1039/B613844F>.
- [48] European Parliament, of the European Union C. Regulation (EU) 2024/1252 of the European parliament and of the Council of 11 April 2024 establishing a framework for ensuring a secure and sustainable supply of critical raw materials and amending regulations (EU) no 168/2013, (EU) 2018/858, (EU) 2018/1724 and (EU) 2019/1020. 2024.
- [49] Pichandi R, Tiwari P, Basu S, Singh S. Ca₂Fe₂O₅ and Ca₂Co₂O₅ based composite cathode for IT-SOFC application. *Solid State Ionics* 2024;404:116418. <https://doi.org/10.1016/j.ssi.2023.116418>.
- [50] Shaula A, Pivak Y, Waerenborgh J, Gaczynski P, Yaremchenko A, Khartov V. Ionic conductivity of brownmillerite-type calcium ferrite under oxidizing conditions. *Solid State Ionics* 2006;177:2923–30. <https://doi.org/10.1016/j.ssi.2006.08.030>.
- [51] Cascos V, Martínez-Coronado R, Alonso JA, Fernández-Díaz MT. Structural and electrical characterization of the Co-doped Ca₂Fe₂O₅ brownmillerite: evaluation as SOFC–cathode materials. *Int J Hydrogen Energy* 2015;40:5456–68. <https://doi.org/10.1016/j.ijhydene.2015.01.067>.
- [52] Squizzato E, Carollo G, Glisenti A. Ca₂Fe_{1.95}Mg_{0.05}O₅: innovative low cost cathode material for intermediate temperature solid oxide fuel cell. *Int J Hydrogen Energy* 2021;46:24555–66. <https://doi.org/10.1016/j.ijhydene.2021.05.020>.
- [53] Li Q, Sun L, Huo L, Zhao H, Grenier J-C. Electrode properties of Co-doped Ca₂Fe₂O₅ as new cathode materials for intermediate-temperature SOFCs. *Int J Hydrogen Energy* 2010;35:9151–7. <https://doi.org/10.1016/j.ijhydene.2010.06.048>.
- [54] Zheng G, Zhang G, Liu K, Huang Z, Chen T, Zhou J, et al. A promising Bi-doped La_{0.8}Sr_{0.2}Ni_{0.2}Fe_{0.8}O_{3-δ} oxygen electrode for reversible solid oxide cells. *Int J Hydrogen Energy* 2023;48:12571–80. <https://doi.org/10.1016/j.ijhydene.2022.12.195>.
- [55] Xu Z, Chang H, Liu C, Tian R, Peng J, Bian L, et al. Bismuth doped La_{0.5-x}BixSr_{0.5}Fe_{0.9}Mo_{0.1}O_{3-δ} perovskite as high active oxygen electrode for solid oxide cells. *Int J Hydrogen Energy* 2024;49:643–9. <https://doi.org/10.1016/j.ijhydene.2024.03.332>.
- [56] Zuo X, Chen Z, Guan C, Chen K, Song S, Xiao G, et al. Molten salt synthesis of high-performance, nanostructured La_{0.6}Sr_{0.4}FeO_{3-δ} oxygen electrode of a reversible solid oxide cell. *Materials* 2020;13:2267. <https://doi.org/10.3390/ma13102267>.
- [57] Wu L, Zheng H, Yang X, Qi H, Tu B, Zang C, et al. Enhanced oxygen evolution rate and Anti-interfacial delamination property of the SrCo_{0.9}Ta_{0.1}O_{3-δ}/La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-δ} oxygen-electrode for solid oxide electrolysis cells. *ACS Appl Mater Interfaces* 2023. <https://doi.org/10.1021/acsami.3c13513>.