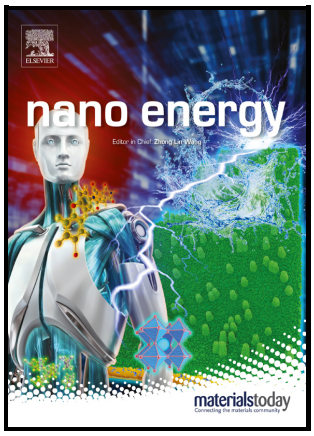
Journal Pre-proof



A General Framework for Quantifying Contact-Electrification of Materials: A Case on the Family of the Canonical Amino Acids

Yijun Hao, Jiyu Su, Xiaopeng Zhu, Yunlei Qie, Jianlong Hong, Xinliang Zhou, Daye Li, Yong Qin, Dong Yang, Wei Su, Hongke Zhang, Yongjiu Yuan, Chengkuo Lee, Zhong Lin Wang, Xiuhan Li

PII: S2211-2855(26)00922-5

DOI: https://doi.org/10.1016/j.nanoen.2026.112618

Reference: NANOEN112618

To appear in: *Nano Energy*

Received date: 29 September 2026
Revised date: 2 November 2026
Accepted date: 12 November 2026

Please cite this article as: Yijun Hao, Jiyu Su, Xiaopeng Zhu, Yunlei Qie, Jianlong Hong, Xinliang Zhou, Daye Li, Yong Qin, Dong Yang, Wei Su, Hongke Zhang, Yongjiu Yuan, Chengkuo Lee, Zhong Lin Wang and Xiuhan Li, A General Framework for Quantifying Contact-Electrification of Materials: A Case on the Family of the Canonical Amino Acids, *Nano Energy,* (2026) doi:https://doi.org/10.1016/j.nanoen.2026.112618

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Journal Pre-proof

A General Framework for Quantifying Contact-Electrification of Materials: A Case on the Family of the Canonical Amino Acids

Yijun Hao[a,c,e], Jiyu Su[a], Xiaopeng Zhu[a], Yunlei Qie[a], Jianlong Hong[c,e], Xinliang Zhou[c,e,g], Daye Li[c,h], Yong Qin[f], Dong Yang[a], Wei Su[a], Hongke Zhang[a], Yongjiu Yuan[d], Chengkuo Lee[c,e,*], Zhong Lin Wang[b,*], Xiuhan Li[a,f,*]

a. School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, 100044, P. R. China.
b. Beijing Institute of Nanoenergy and Nanosystems, Chinese Academy of Sciences, Beijing 101400, P. R. China.
c. Department of Electrical and Computer Engineering, National University of Singapore, 117576, Singapore.
d. Laser Micro/Nano-Fabrication Laboratory, School of Mechanical Engineering, Beijing Institute of Technology, Beijing, 100081, China.
e. Center for Intelligent Sensors and MEMS (CISM), National University of Singapore, 117583. Singapore.
f. State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing, 100044, P.R. China.
g. State Key Laboratory of Advanced Fiber Materials, College of Materials Science and Engineering, Donghua University, Shanghai, 201620, China.
h. School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, 100083, China.

ORCID: Yijun Hao, 0009-0009-3146-1903; Xiuhan Li: 0000-0002-5254-9877.
E-mail: Chengkuo Lee, elelc@nus.edu.sg; Zhong Lin Wang, wangzhonglin@binn.cas.cn; Xiuhan Li, lixiuhan@bjtu.edu.cn.

Abstract

Contact-electrification (CE) is the earliest phenomenon recorded in human civilization, but its understanding remains a challenge till now. Materials for CE have largely been developed through empirical screening, with their performance commonly assessed using electrical output and triboelectric ranking. Although these approaches have enabled substantial advances in triboelectric energy harvesting, they provide limited quantitative insight into intrinsic CE capability and the molecular differences underlying material performance. Here we establish an interpretable framework linking molecular electronic descriptors to experimentally measured CE responses in a common platform by integrating systematic experiments, quantum-chemical analysis and interpretable learning. The framework reconstructs a continuous molecular landscape of CE across all twenty canonical amino acids, as an example, reveals that interfacial charge generation is governed by cooperative global and local electronic characteristics, and enables framework-guided molecular discovery with experimentally validated enhancement in both CE capability and energy-storage performance. By quantitatively bridging molecular electronic characteristics with CE capability and electrical output, this work transforms empirical output comparison and material ranking into descriptor-guided molecular design, enabling

rational optimization of mechanical-to-electrical transduction for high-performance energy harvester.

Journal Pre-proof

Main

Contact-electrification (CE), the spontaneous generation of electrical charges at material interfaces,[1,2,3] is a fundamental physics phenomenon known for >2600 years. CE has been utilized for an energy-conversion process that enables the direct harvesting of ubiquitous mechanical energy through triboelectric nanogenerators (TENGs).[4,5,6,7] During contact between two materials, the tendency of a material to gain or lose electrons is governed by its intrinsic electronic characteristics, whereas the polarity of the generated charges is determined by its relative electron-attracting capability with respect to the contacting counterpart.[8,9,10] Continuous advances in material engineering have substantially enhanced CE sensing performance, accelerating the development of sustainable self-powered technologies for wearable electronics,[11,12,13] distributed sensors,[14,15,16] intelligent infrastructures[17,18] and self-powered systems.[19,20,21,22] Despite these achievements, the discovery and optimization of triboelectric materials remain largely empirical, relying on iterative experimental screening and output-oriented evaluation.

Current assessment of triboelectric materials is predominantly based on TENG electrical output, which also serves as the sensing signal in self-powered sensors.[23,24,25] This output-oriented approach has produced triboelectric series that rank materials according to their relative charging tendencies and has been useful for identifying high-performance material pairs.[26,27,28,29] However, such rankings mainly provide order information. They indicate which material is more positive or negative, but do not quantify the intrinsic magnitude of CE capability or the molecular difference between adjacent materials in the series. As a result, a material with higher output is often regarded as having stronger CE capability, without a quantitative definition of what this capability represents at the molecular level. This limitation restricts the transition from material comparison to molecular design. Without an intrinsic molecular metric, CE behavior cannot be directly connected to molecular electronic characteristics, and subtle structural variations cannot be translated into predictable differences in charge generation. Consequently, the optimization of triboelectric materials still depends largely on empirical screening and repeated device-level testing. An interpretable framework that quantitatively defines CE capability and links molecular structure with interfacial charge generation is therefore required.

Establishing such a framework requires a molecular library that simultaneously minimizes structural complexity and preserves systematic diversity. Canonical amino acids uniquely satisfy these criteria by sharing a conserved backbone while differing only in distinct side chains (SCs), thereby providing an ideal molecular space for isolating the electronic origins of CE. Here we establish an interpretable framework for CE by integrating systematic molecular experiments, quantum-chemical descriptors and interpretable machine learning based on all canonical amino acids. The resulting framework quantitatively maps CE capability on a continuous molecular scale, identifies the key electronic characteristics governing interfacial charge generation and enables framework-guided molecular discovery beyond naturally occurring amino acids. By quantitatively bridging molecular electronic descriptors and CE electrical output, this work transforms empirical sensing material optimization into descriptor-guided design, establishing a molecular foundation for mechanical-to-electrical transduction.

A Molecular Library of Canonical Amino Acids

Establishing an interpretable molecular framework requires a molecular library that minimizes structural complexity while preserving systematic chemical diversity. Canonical amino acids satisfy these criteria by sharing a conserved molecular backbone and differing only in their SC structures, thereby forming one of the most structurally ordered yet chemically diverse molecular systems in nature. Such a common structural scaffold enables subtle molecular variations to be examined without introducing excessive structural complexity, making amino acids well suited for investigating the molecular origins of CE. Moreover, as the molecular precursors of numerous naturally occurring and synthetic derivatives, the twenty canonical amino acids define a compact yet representative chemical space with broad physicochemical coverage. Compared with conventional material libraries, where molecular backbones, functional groups and structural topologies often vary simultaneously, amino acids provide a highly controlled system in which electronic differences can be directly associated with specific structural modifications. This characteristic enables molecular-level structure–property relationships to be examined with minimal interference from unrelated structural factors.

All twenty canonical amino acids were therefore selected to construct the molecular library (Fig. 1a). Based on the physicochemical characteristics of their SCs, they can be classified into four representative groups: polar (ARG, HIS, LYS, ASP and GLU), nonpolar (SER, THR, ASN, GLN, CYS and GLY), aliphatic (ALA, VAL, ILE, LEU, MET and PRO) and aromatic (PHE, TYR and TRP). Together, these groups span a broad range of charge states, hydrogen-bonding capabilities, hydrophobicity and π-conjugation, generating a chemically diverse yet structurally coherent molecular space. Because all molecular variations remain confined to the SCs while the molecular backbone remains unchanged, differences in electronic behavior can be directly linked to well-defined structural modifications. The resulting library therefore combines broad chemical diversity with strict structural consistency, providing a suitable platform for comparing molecular electronic characteristics across different amino acid families. Figure 1b and Supplementary Fig. 1 further summarize the SC functional characteristics underlying this classification and illustrate the molecular diversity contained within the amino acid library.

Although all amino acids retain the same molecular backbone, their chemically distinct SCs introduce systematic differences in local electron distribution, polarity, frontier-orbital characteristics and site-specific reactivity (Fig. 1c). These electronic variations define the intrinsic characteristics of individual amino acids while maintaining a structurally clean and highly interpretable molecular system. The magnitude of these differences is sufficiently large to generate distinguishable molecular behaviors, yet sufficiently constrained to preserve meaningful comparisons throughout the entire library. As a result, the amino acid library does not consist of isolated molecular classes but instead defines a continuous and chemically connected molecular space. This combination of structural uniformity, chemical diversity and electronic distinguishability provides the basis for quantitatively mapping CE capability, identifying its molecular origins, establishing continuous structure–property relationships and guiding subsequent framework-driven molecular discovery.

Establishing the Molecular Landscape

CE arises from interfacial charge transfer between two dissimilar materials during repeated contact and separation processes (Supplementary Fig. 2). Quantitative comparison of molecular CE capability requires a device architecture in which molecular variation remains the only significant experimental variable. Poly(vinylidene fluoride) (PVDF) was selected as the host matrix because of its excellent film-forming ability, chemical stability and widespread use in triboelectric systems. Amino acids were incorporated into PVDF to form amino acid-composite films (AF), while pristine PVDF films (PF) served as the counterpart triboelectric layer (Fig. 2a–b). Since both triboelectric layers share the same polymer matrix, differences in electrical output primarily originate from the molecular characteristics introduced by the amino acid SCs. Uniform films were prepared using a two-step spin-coating process (Supplementary Table 1). The resulting AFs remain continuous during bending and rolling without visible cracking or delamination, demonstrating excellent flexibility and mechanical integrity (Fig. 2c). Surface SEM images show smooth and homogeneous morphologies for both PF and AF, with no obvious structural differences observed at the microscale (Supplementary Fig. 3). Cross-sectional SEM images reveal intimate contact between the AF layer and the Al substrate across multiple length scales (Fig. 2d). Elemental mapping further confirms homogeneous amino acid incorporation throughout the composite films, where the uniformly distributed nitrogen signal distinguishes AF from PF (Fig. 2e and Supplementary Fig. 5). These observations indicate that structural and morphological variations remain minimal, allowing subsequent electrical measurements.

Under identical fabrication and testing conditions, the twenty canonical amino acids span a broad molecular landscape of CE capability. The measured open-circuit voltage increases progressively from Asp (0.1247) and Glu (0.1551) through CYS, MET, PHE, ASN, TRP, GLN, THR, TYR, SER, GLY, ILE, VAL, LEU, ALA and PRO to HIS (0.8351), LYS (0.9042) and ARG (0.9322) (Fig. 2f). Rather than separating into several discrete groups, the amino acids occupy a nearly continuous distribution across the entire capability range, indicating that relatively small changes in SC chemistry produce measurable differences in interfacial charge generation. This experimentally established landscape provides a quantitative description of CE capability at the molecular level. Pairwise contact measurements between representative AFs were subsequently carried out to examine the internal consistency of the landscape (Fig. 2g). When two AFs with different CE capabilities undergo contact and separation, the initial current polarity consistently points from the lower-capability material toward the higher-capability counterpart, providing a direct comparison independent of the PVDF reference layer. Across all tested pairs, the measured current directions agree with the experimentally established ordering, while the current magnitudes follow the relative capability differences between neighboring amino acids. The pairwise measurements therefore reproduce both the direction and magnitude trends observed in the molecular landscape, supporting the validity of the experimentally determined CE capability distribution across the twenty canonical amino acids. To assess whether changes in the PVDF crystalline phase could contribute to differences in measured CE output, FTIR spectra were collected from PVDF and films containing representative amino acids (Supplementary Fig. 6). All examined films retain the characteristic α-phase bands near 614, 763, 795 and 975 cm^{-1}, with no distinct β-phase band emerging near 1275 cm^{-1}. These results indicate that the measured output differences are not accompanied by clear FTIR evidence of

pronounced α-to-β conversion, supporting the preservation of α-phase structural features following amino-acid incorporation.

Descriptor Representation of Molecular Characteristics

Transforming the experimentally established CE landscape into a learnable molecular space requires a representation that preserves physical meaning while remaining sufficiently compact for quantitative analysis. Rather than maximizing descriptor quantity, only descriptors satisfying chemical interpretability, transferability and statistical robustness were retained (Supplementary Note 1). Because natural amino acids share an identical molecular backbone and differ only in their SC structures (Supplementary Fig. 7), molecular characteristics can be naturally described from two complementary perspectives: global electronic characteristics and local reactive tendencies (Fig. 3a). The global representation consists of electron affinity (EA), electron donation capability (ED) and energy gap (GAP), which describe frontier-orbital characteristics governing long-range electron redistribution during CE. Conventional CE interpretations often associate stronger electron-accepting capability with enhanced charge-transfer tendencies, while the combination of higher electron affinity and a smaller energy gap is generally considered more favorable for interfacial charge generation (Supplementary Fig. 8). The local representation is derived from condensed Fukui functions by separately evaluating the amino group, carboxyl group and SC under electron-accepting and electron-donating conditions, yielding six descriptors: amino nucleophilicity (AN), amino electrophilicity (AE), carboxyl nucleophilicity (CBN), carboxyl electrophilicity (CBE), side-chain nucleophilicity (SN) and side-chain electrophilicity (SE). Together, these descriptors capture both long-range electronic characteristics and site-specific charge-transfer tendencies, providing a physically interpretable representation of molecular CE behavior.

Density functional theory calculations were subsequently performed using the parameters summarized in Supplementary Table 2. Optimized molecular geometries, frontier-orbital distributions and the complete descriptor matrix are provided in Supplementary Table. 3–22 and Supplementary Table 23. Despite differing only in SC structures, the twenty amino acids span continuous ranges across every descriptor (Fig. 3b and Supplementary Fig. 9-13), indicating that relatively small structural modifications generate measurable electronic diversity. For example, EA varies from −0.07195 to −0.03748, while local descriptors such as CBE and SN span nearly an order of magnitude across the molecular library. Similar continuous distributions are observed for ED, GAP and all Fukui-derived descriptors, demonstrating that the amino acid library does not comprise several discrete electronic classes but instead defines a broad continuum of molecular electronic characteristics. Following normalization and semantic transformation (Fig. 3c–d), all descriptors are mapped into a unified descriptor space that enables direct comparison of electronic and reactive contributions. The calculated descriptors provide a common reference for representing molecular differences across the library. A mapping between these descriptors and the CE responses measured in the AF-PF platform is built. Pearson analysis reveals that no individual descriptor dominates CE behavior (Fig. 3e and Supplementary Fig. 14). EA exhibits the strongest positive correlation with experimentally measured CE capability, followed by CBE, whereas SN displays a weak linear associations. Several descriptors, including AN, AE and ED, exhibit only weak correlations despite their clear chemical relevance. Even the strongest individual correlation explains only

part of the observed molecular variation, indicating that CE cannot be reduced to a single electronic parameter. Instead, the measured capability emerges from coupled contributions involving multiple electronic and reactive characteristics. This observation further suggests that the experimentally established CE landscape is controlled by descriptor interactions rather than descriptor magnitude alone, highlighting the limitations of conventional single-parameter interpretations frequently adopted in triboelectric material studies.

The collective organization of these descriptors provides additional insight into the structure of the molecular space. Projection into a two-dimensional descriptor embedding reveals that global and local descriptors occupy distinct yet complementary regions rather than collapsing into a single cluster (Fig. 3f). Their spatial separation indicates that each descriptor group contributes nonredundant information to molecular representation. Distances from the embedding center further quantify their relative influence on the overall descriptor organization (Fig. 3g and Supplementary Fig. 15–16). SE (0.575), CBE (0.571) and GAP (0.491) occupy the most distant positions, whereas descriptors such as AN (0.389) and AE (0.413) remain closer to the embedding center. This distribution suggests that SC reactivity, carboxyl-group electrophilicity and frontier-orbital characteristics contribute most strongly to descriptor diversity across the amino acid library. Notably, descriptors exhibiting the largest embedding distances are not necessarily those displaying the strongest individual Pearson correlations, further supporting the notion that molecular CE behavior arises from collective descriptor organization rather than isolated descriptor effects.

When the natural amino acids are projected into the same embedding, they populate a continuous molecular landscape rather than forming isolated clusters (Fig. 3h). Amino acids belonging to different physicochemical classes remain partially interconnected despite their discrete chemical classifications, indicating that molecular electronic characteristics evolve continuously across the library. For example, ARG, LYS and HIS occupy a distinct region associated with high CE capability, whereas ASP and GLU are positioned toward the opposite side of the embedding. Nevertheless, intermediate amino acids bridge these regions through gradual descriptor transitions rather than abrupt boundaries. The embedding can be further partitioned into four descriptor regions according to the signs of the two coordinates (Fig. 3i), each corresponding to distinct combinations of global electronic characteristics and local reactive tendencies. The absence of sharp boundaries between neighboring regions suggests that molecular evolution within the amino acid space proceeds through continuous descriptor variation rather than discrete class-to-class transitions. Consequently, the amino acid library defines a quantitatively navigable molecular space in which discrete molecular identities are transformed into continuous electronic coordinates. This representation provides the basis for learning structure-property relationships, identifying governing molecular characteristics and enabling subsequent framework-guided molecular discovery.

Journal Pre-proof

Quantifying an Interpretable Molecular Framework

Although the descriptor space captures systematic electronic variations across the amino acid library, individual descriptors exhibit only weak-to-moderate linear relationships with experimentally measured CE capability. Instead, the descriptor matrix reveals extensive coupling among electronic and reactive representations (Fig. 4a). Global electronic descriptors display moderate positive correlations with carboxyl-related reactivity, whereas SC descriptors exhibit pronounced negative correlations with both electronic characteristics and amino-group reactivity. Strong anti-correlations are also observed between SC electrophilicity and carboxyl electrophilicity, indicating that local reactive tendencies are not independent but evolve cooperatively within the molecular space. Such relationships suggest that CE emerges from coordinated molecular electronic effects rather than independent descriptor contributions. The experimentally established CE landscape therefore represents a multivariate molecular phenomenon in which multiple electronic characteristics determine charge-generation behavior. This descriptor coupling further implies that structure–property relationships are unlikely to be captured through conventional one-descriptor-at-a-time interpretations.

To determine whether these coupled relationships can be quantitatively learned, a support vector regression framework was constructed using the descriptor representation. As shown in Fig. 4b, the reconstructed CE landscape closely reproduces both the overall ranking and the local differences among canonical amino acids. Amino acids with high CE, including ARG, LYS and HIS, remain positioned at the upper end of the reconstructed landscape, whereas ASP, GLU and CYS consistently occupy the lower region. Intermediate amino acids also preserve their relative ordering, indicating that the framework captures subtle molecular variations. More importantly, the framework preserves the continuous organization of CE capability rather than reducing molecules into discrete categories, suggesting that descriptor evolution remains consistent with the experimentally observed molecular landscape. Comparison with additional learning algorithms highlights the learnability of the molecular space (Fig. 4c). Support vector regression achieves a LOOCV R^2 of 0.8089 together with a MAE of 0.0850 and RMSE of 0.1001, substantially outperforming Gaussian process regression (R^2 = 0.2084), ridge regression (R^2 = 0.1369) and linear regression (R^2 = 0.0077). Hyperparameters are listed in Supplementary Table 24. Although Gaussian process regression reproduces the training data, its poor LOOCV performance indicates severe overfitting, whereas the near-zero predictive capability of linear regression suggests that CE cannot be described through simple additive descriptor effects. These observations indicate that CE follows nonlinear but learnable molecular rules embedded within the descriptor space.

Framework performance further depends on descriptor representation. Five candidate descriptor representations ranging from complete molecular representations to progressively simplified subsets were therefore evaluated (Fig. 4d). Increasing descriptor dimensionality does not continuously improve framework performance. Instead, the highest predictive capability is achieved using only five descriptors, EA, GAP, CBE, SN and SE. The resulting descriptor hierarchy exhibits a clear performance gradient, with LOOCV R^2 decreasing from 0.8089 to 0.3978 (Fig. 4e(i)). Descriptor contribution analysis reveals that CBE contributes most strongly to framework learning, followed by EA, SN and GAP, whereas SE acts primarily as a complementary descriptor (Fig. 4e(ii)). Although the framework provides an interpretable description of descriptor–response relationships, a high model-assigned importance does not

necessarily identify the dominant physical factor governing CE. These importance estimates depend on the size and quality of the dataset and should therefore be interpreted within the scope of the present measurements. Remarkably, descriptors originating from both global electronic characteristics and local reactive tendencies are retained within the optimal representation, indicating that neither category alone is sufficient to describe CE behavior. A molecular landscape spanning twenty chemically distinct amino acids can therefore be reconstructed using only five physically interpretable descriptors. Comparison with recent descriptor-driven materials-learning studies further places the present framework within an underexplored regime of extremely small datasets and low descriptor dimensionality (Fig. 4f and Supplementary Table 25). While most existing frameworks rely on hundreds of samples and tens of descriptors to achieve high predictive accuracy, the present approach extracts transferable molecular rules from only twenty molecules using a compact descriptor representation.[30,31,32,33,34,35] The framework therefore emphasizes mechanistic interpretability and molecular rule extraction. The framework also provides direct insight into how molecular characteristics regulate CE capability. Descriptor–response profiles reveal two distinct model response patterns (Fig. 4g). CBE and SN exhibit nearly monotonic enhancement of CE capability with increasing descriptor values, indicating persistent positive contributions to interfacial charge generation. In contrast, EA, GAP and SE display pronounced optimal-response windows. Increasing these descriptors initially enhances CE capability, followed by a gradual reduction beyond a characteristic range. Such behavior suggests that stronger electronic characteristics do not necessarily translate into stronger CE performance and that efficient charge generation requires a balance among multiple molecular factors. The framework therefore reveals that CE optimization is governed not by maximizing individual descriptors but by locating favorable regions within a multidimensional electronic landscape. Residual analysis provides additional insight into framework generalization (Supplementary Fig. 17). Aliphatic amino acids exhibit the smallest prediction deviations, indicating that their relatively simple SC structures are well captured by the descriptor representation. Polar amino acids also maintain uniformly distributed residuals without systematic overestimation or underestimation. These residual patterns suggest that the descriptor representation captures some structural variations more consistently than others, motivating further assessment of how predictive performance depends on model selection and chemical coverage.

The non-nested LOOCV analysis above was initially used to explore compact, physically interpretable descriptor combinations and their relationships with the CE landscape. To evaluate the complete model selection procedure, nested LOOCV was subsequently performed, with descriptor selection and hyperparameter optimization restricted to the inner loop and standardization fitted within each training fold (Fig. 4h). The pooled outer predictions yielded an R^2 of 0.6333 across all twenty amino acids, providing an assessment that incorporates variability associated with model selection. In this small library, CYS, MET and TYR are the sole representatives of thiol, thioether and phenolic side-chain functionalities, respectively. When one is withheld, its corresponding functionality is absent from the training set. To examine sensitivity to this sparse chemical coverage, these three amino acids were omitted in a post hoc targeted exclusion analysis. Repeating nested LOOCV on the remaining seventeen molecules yielded an R^2 of 0.8962 using EA, GAP, CBE and SN. With the original five descriptors held fixed and only hyperparameters optimized through inner LOOCV, the

framework still achieved an R^2 of 0.8712 on the same subset (Fig. 4i). Repeating the original non-nested descriptor and hyperparameter search on the same seventeen-molecule subset yielded an R^2 of 0.8992, again selecting EA, GAP, CBE and SN. Retaining the original five descriptors and performing only non-nested hyperparameter optimization gave an R^2 of 0.8629. These values characterize a retrospectively defined subset and are interpreted within its chemical coverage. The strong performance obtained without changing the original descriptor combination supports its continued use as a compact, physically interpretable representation for prediction within the retained molecular space.

The framework converts the experimentally established CE landscape into a quantitatively learnable structure–property relationship. Rather than predicting molecular performance from empirical ranking alone, it identifies electronic descriptor patterns associated with CE capability, reveals their cooperative regulation modes and transforms molecular CE behavior into a navigable descriptor space. These molecular rules explain the experimentally observed landscape and provide the basis for subsequent framework-guided molecular discovery.

Framework-Guided Molecular Discovery

The molecular framework established above provides an opportunity to explore CE capability beyond the twenty canonical amino acids. Because the purpose is to evaluate framework-guided molecular evolution rather than arbitrary molecular modification, a structurally conservative derivatization strategy was adopted. Among the many naturally occurring and synthetic amino acid derivatives, SC extension is one of the most common modification routes and introduces minimal perturbation to the original molecular identity. By preserving the parent functional groups while gradually increasing the separation between the backbone and SC, this strategy enables continuous tuning of molecular electronic characteristics without introducing additional chemical functionalities. Therefore, one to three methylene (-CH_2-) units were inserted between the amino acid backbone and SC (Step1–3), generating a derivative library that remains chemically connected to the original amino acid space while extending the accessible molecular landscape. All derivatives were evaluated using the same descriptor construction procedure and projected into the optimal descriptor space established above. Bootstrap resampling (500 iterations) was performed to estimate prediction uncertainty. The resulting evolutionary trajectories exhibit markedly different behaviors (Fig. 5a(i–vi) and Supplementary Fig. 18–20). Positively charged amino acids generally retain high predicted outputs throughout derivatization, whereas several polar and aromatic amino acids show pronounced non-monotonic responses. In some cases, CE capability increases continuously with chain extension; in others, it decreases or fluctuates substantially. These observations indicate that molecular evolution cannot be described by geometric modification alone and remains strongly dependent on the intrinsic electronic characteristics of the parent amino acid. Ranking derivatives according to predicted output identifies several high-performing candidates, including CYS Step2, HIS Step3, HIS Step1, ARG Step2 and ARG Step1 (Fig. 5b). However, molecules with similar predicted outputs often exhibit substantially different bootstrap uncertainties, suggesting that predicted performance alone is insufficient for prioritizing experimental validation.

To account for both performance and prediction reliability, a Prediction Priority Score (PPS) was introduced by integrating predicted output, uncertainty and molecular evolutionary distance into

a single metric (Fig. 5c). The resulting ranking differs significantly from the raw output ranking. Molecules such as ARG Step1, HIS Step1 and GLN Step1 move to the top of the candidate list, whereas several high-output molecules with large uncertainties are deprioritized (Fig. 5d). Analysis of the relationships among output, PPS, uncertainty and molecular evolution reveals a clear trend (Fig. 5e(i–iv) and Supplementary Fig. 21): molecules with higher PPS values generally exhibit lower uncertainty, whereas larger evolutionary distances are frequently accompanied by reduced prediction confidence. This pattern suggests that the framework remains most reliable within regions that remain electronically connected to the original amino acid descriptor space. When derivatives are grouped according to SC chemistry, several transferable trends emerge (Fig. 5f(i–iv)). Guanidinium-containing and primary-amine SCs consistently retain the highest predicted CE capability following derivatization, whereas sulfur-containing and aromatic SCs display substantially greater variability. Notably, the evolutionary hierarchy largely follows the ordering observed in the parent amino acid landscape, indicating that the framework captures persistent molecular characteristics rather than derivative-specific behavior.

ARG Step1 and GLN Step1 were selected for experimental validation based on their predicted outputs, PPS rankings, molecular representativeness and commercial availability (Fig. 5g(i,ii)). FTIR spectra further confirm the expected molecular evolution (Supplementary Fig. 22). Both derivatives display strengthened absorption bands near 2920 and 2850 cm^{-1} relative to PF, characteristic of CH_2 stretching vibrations introduced by side-chain extension, consistent with the designed molecular structures. These two derivatives represent distinct scenarios. ARG Step1 maintains a CE capability comparable to that of its parent molecule while achieving the highest PPS in the derivative library. In contrast, GLN Step1 is predicted to increase from 0.46 to 0.74, corresponding to an improvement of approximately 60%. Experimental measurements closely match the predicted values, yielding 0.88 for ARG Step1 and 0.78 for GLN Step1 (Fig. 5h(i–iii)). The agreement extends beyond ranking and reproduces the magnitude of molecular evolution relative to the parent amino acids, supporting both framework accuracy and practical utility. Beyond increasing electrical output, the enhanced CE capability directly improves the quality of self-powered sensing signals. Because the sensing response originates from CE, a higher electrical output increases the separation between the mechanically induced signal and the background electrical noise. Under otherwise identical device and measurement conditions, this enhancement improves the signal-to-noise ratio (SNR) without external amplification or additional signal-processing circuitry. Representative current traces show that GLN Step1 produces a substantially stronger and more readily distinguishable response than its parent GLN under the same mechanical stimulation (Supplementary Fig. 23). The enhanced current amplitude increases the SNR from 14.9 to 22.9 dB (Fig. 5i), demonstrating that framework-guided molecular evolution improves both signal distinguishability and measurement reliability.

The molecular-level enhancement in CE capability also translates directly into energy-storage performance. Capacitor-charging measurements show that derivatives with higher CE capability achieve faster charging and greater stored energy, while the experimentally measured charging curves closely follow the charging behaviour reconstructed from the framework (Fig. 5j). These results connect molecular electronic characteristics, CE capability, electrical output, sensing quality and energy accumulation within a unified framework. The identified molecular rules therefore remain predictive across multiple levels of CE-based signal

generation and mechanical-energy conversion. Although the present framework is not intended as a universal predictor for arbitrary molecular systems, it establishes a descriptor-guided strategy for extracting transferable molecular design principles and translating them into both sensor-level performance and practical energy-storage capability.

Conclusions and outlook

Using twenty canonical amino acids as a structurally related molecular library, this work establishes a framework linking molecular electronic characteristics with experimentally measured contact-electrification (CE) capability. A compact representation combining global electronic and local reactivity descriptors captures nonlinear relationships within the measured CE landscape. Integrating predicted output, prediction uncertainty and descriptor-space distance through the Prediction Priority Score (PPS) further supports the prioritization of related derivatives. Experimental tests of 2 amnio acid-derivatives illustrate prediction consistency and performance enhancement, respectively, providing initial experimental support for the descriptor-guided discovery strategy. However, small dataset forms a chemically defined and inherently library with uneven representation of side-chain functionalities. Experimental validation focused on two derivatives prioritized by prediction confidence and compound availability. Physically meaningful descriptors support interpretation of the learned relationships, although the statistical model does not explicitly resolve atom states. Extension to other molecular classes, film compositions and operating conditions requires further validation. Within this defined scope, the study provides a worked example of connecting molecular representation, quantitative learning and experimental evaluation, offering a methodological reference for broader CE materials research.

Future development should combine learning from related molecular data with systematic expansion of reliable CE measurements (Supplementary Note 3). Broader structural diversity would allow the identified relationships to be examined more thoroughly, while studies across different material environments would assess their transferability. Experiments should both test model predictions and provide new observations for refining the molecular representation. Connecting these activities could build a cumulative basis for CE materials design, in which each prediction generates a testable molecular hypothesis and each experiment strengthens the knowledge available for subsequent discovery. This approach could support more systematic development of self-powered materials and mechanical-to-electrical transduction technologies.

Journal Pre-proof

Methods

Materials:

All amino acids (>99%), Poly(vinylidene fluoride) powder (average Mw: 400000), and dimethylformamide (99.5%) were purchased from Macklin.

Film fabrication:

Pristine PVDF films and amino acid-composite PVDF films were prepared using an identical spin-coating procedure. The precursor solutions were first magnetically stirred for 2 h to ensure complete dissolution and homogeneous dispersion. Subsequently, the solution was dropped onto the center of an Al substrate and deposited through a two-step spin-coating process consisting of an initial low-speed spreading step (600 rpm, 15 s) followed by a high-speed coating step (1800 rpm, 30 s). The resulting films were thermally annealed at 80 °C for 30 min and naturally cooled to room temperature. This process produced uniform free-standing films suitable for subsequent measurements.

Morphological, elemental, electrical characterizations:

Surface morphologies of PF and AF were characterized using field-emission scanning electron microscopy (FE-SEM, ZEISS GeminiSEM 300, Germany). Cross-sectional morphologies were obtained by fracturing the films and observing the exposed interfaces under identical imaging conditions. Elemental distributions were analyzed by energy-dispersive X-ray spectroscopy (EDS) coupled with SEM. Fluorine elemental mapping was used to visualize the PVDF matrix, whereas nitrogen elemental mapping was employed as a characteristic indicator of amino acid incorporation. Optical images of the free-standing films were recorded under identical illumination conditions to evaluate their mechanical integrity and film uniformity. The chemical structures of the films were characterized by Fourier transform infrared spectroscopy (FTIR, Thermo Scientific Nicolet iS20, USA). Spectra were collected over the range of 4000–400 cm^{-1}. Electrical outputs were measured using a triboelectric testing platform consisting of an optical breadboard, a programmable linear motor and an electrometer. Periodic contact and separation motions were driven by the linear motor (J-Best RP-200), while the generated open-circuit voltage, short-circuit current were recorded in real time using a programmable electrometer (Keithley 6514). Unless otherwise specified, all electrical measurements were performed using identical device configurations and operating conditions for direct comparison among different amino acid-composite films.

Contact-electrification (CE) measurements:

The amino acid-composite PVDF film (AF) and pristine PVDF film (PF) were assembled as the triboelectric pair for the primary CE measurements. Unless otherwise specified, all amino acid films were fabricated using identical procedures and measured under identical experimental conditions. The experimentally measured open-circuit voltage was adopted as a quantitative indicator of relative CE response under these conditions. For each amino acid formulation, four independently fabricated films from 4 preparation batches were tested. Reproducible voltage responses were obtained across the independently prepared films and batches. To further examine the consistency of the experimentally established molecular landscape, pairwise CE measurements were performed between representative AF. The polarity of the initial current pulse was used to assess the relative charging order and its consistency with the voltage-based ranking.

CE testing environment:

Unless otherwise specified, all CE measurements were performed under identical laboratory conditions to enable direct comparison among different amino acid formulations. The ambient temperature was maintained at 25 ± 0.5 °C, and the relative humidity was controlled at 40%. All samples were equilibrated under these conditions for at least 30 min before measurement. Contact–separation measurements were performed at a contact force of 1 N, a frequency of 1 Hz and a separation distance of 20 mm, with an effective contact area of 1 cm^2. The same device configuration and measurement system were used throughout the sample library. For each type of electrical measurement, acquisition parameters were kept unchanged across samples.

Energy storage measurements:

Capacitor-charging measurements were performed to evaluate the practical energy-storage capability enabled by molecular CE. AF and PF served as the triboelectric pair, with aluminium foils acting as the electrodes for both layers. The generated alternating electrical output was converted into direct current through a full-wave bridge rectifier (DB107) and subsequently stored in a commercial capacitor with a capacitance of 1.0 μF. Periodic contact–separation motion was actuated by J-Best RP-200 operating at a frequency of 1 Hz, an applied force of 1 N, and a separation distance of 20 mm.

Quantum-chemical calculations:

Quantum-chemical calculations were performed through a three-step workflow consisting of classical geometry optimization, density functional theory optimization and electronic structure calculations. Initial molecular geometries were first optimized using the Forcite module with the Universal force field under ultra-fine quality settings. Atomic charges were assigned using the current charge scheme, while both electrostatic and van der Waals interactions were treated using atom-based summation methods. The pre-optimized structures were subsequently refined by geometry optimization using the $DMol^3$ package with the GGA-PBE exchange-correlation functional under spin-restricted conditions. Neutral molecular configurations (charge = 0) were employed with symmetry constraints enabled. Geometry optimization was performed with a maximum of 500 iterations using convergence criteria of 1.0×10^{-5} Ha for energy, 0.002 Ha $Å^{-1}$ for force and 0.005 Å for atomic displacement. Electronic structure calculations were then carried out using the optimized geometries within the $DMol^3$ Energy module. All-electron calculations were performed using the GGA-PBE functional together with the DND basis set under fine orbital cutoff quality. Self-consistent field calculations employed a convergence tolerance of 1.0×10^{-6} Ha with a maximum of 500 SCF cycles. Multipolar expansion was treated to the hexadecapole level, and direct inversion in the iterative subspace (DIIS) acceleration was enabled throughout all calculations.

Quantum-chemical calculations and descriptor construction

Global electronic descriptors were obtained from the calculated frontier orbital energies. The lowest unoccupied molecular orbital (LUMO) energy was used as the electron-accepting descriptor (EA), while the highest occupied molecular orbital (HOMO) energy was used as the electron-donating descriptor (ED). The energy gap (GAP) was calculated as the difference between the LUMO and HOMO energies. Local electronic characteristics were represented using condensed Fukui indices derived from atomic charge calculation. $Fukui^+$ describes the local electronic response to electron addition and characterizes electron-accepting tendencies, whereas $Fukui^-$ describes the response to electron removal and characterizes electron-

donating tendencies. The corresponding atomic indices were summed over the atoms assigned to the amino group, carboxyl group and side chain. The group-resolved Fukui$^-$ values were denoted amino nucleophilicity (AN), carboxyl nucleophilicity (CBN) and side-chain nucleophilicity (SN), while the corresponding Fukui$^+$ values were denoted amino electrophilicity (AE), carboxyl electrophilicity (CBE) and side-chain electrophilicity (SE). These global and local descriptors were subsequently used to construct the molecular representations for the learning analyses.

Statistical analysis:

All experimental measurements were independently repeated ten times unless otherwise specified, and averaged values were used for quantitative analysis. Descriptor contributions were evaluated using permutation importance analysis, whereas descriptor–response relationships were interpreted using partial dependence analysis. Pairwise descriptor relationships were analyzed using Pearson correlation coefficients. All machine-learning analyses were implemented in Python, and all quantum-chemical calculations were performed using BIOVIA Materials Studio.

Experimental reproducibility:

Consistent fabrication procedures and electrical measurement conditions were applied throughout the amino acid library to minimize procedural variability and improve comparability. Reproducible voltage responses were obtained across independently fabricated films from different preparation batches. The measured open-circuit voltage was therefore used as an operational indicator of relative CE response within the AF–PF platform, supporting descriptor–response analysis under the specified conditions.

Data Availability

The data that support the findings of this study are available from the corresponding author upon request.

Code Availability

Selected Python code is publicly available at https://github.com/Yi-Jun16/Amino-acid-s-parameter-CE-mapping.git.

References

1. Li, H., et al. A contact-electro-catalytic cathode recycling method for spent lithium-ion batteries. *Nat. Energy* **8**, 1137–1144 (2023).

2. Sobarzo, J.C., et al. Spontaneous ordering of identical materials into a triboelectric series. *Nature* **638**, 664–669 (2025).

3. Grosjean, G., et al. Adventitious carbon breaks symmetry in oxide contact electrification. *Nature* **651**, 626–631 (2026).

4. Hinchet, R., et al. Transcutaneous ultrasound energy harvesting using capacitive triboelectric technology. *Science* **365**, 491–494 (2019).

5. Xu, W., et al. A droplet-based electricity generator with high instantaneous power density. *Nature* **578**, 392–396 (2020).

6. Chen, J., et al. Reviving Vibration Energy Harvesting and Self-Powered Sensing by a Triboelectric Nanogenerator. *Joule* **1**, 480–521 (2017).

7. Pal, A., et al. Contact Electrification–Based Enantioselective Recognition of Chiral Amino Acids through Stereospecific Interfacial Electron Transfer. *J. Am. Chem. Soc.* **148**, 3, 3063–3076 (2026).

8. Li, S., et al. Biomimetic multimodal tactile sensing enables human-like robotic perception. *Nat. Sens.* **1**, 52–62 (2026).
9. Hao, Y., et al. A Compact Phase-Aligned Hybrid Energy Harvester for Efficient Low-Frequency Energy Harvesting. *Adv. Funct. Mater.* **36**, e74920 (2026).
10. Hao, Y., et al. Contact‐Electrification‐Regulated Amino Acid for Improving Materials' Triboelectric Properties. *Adv. Mater.* **38**, e74012 (2026).
11. Hong, K., et al. Sodium-Alginate Composite Nanofiber-Based Triboelectric Sensor for Self-Powered Wrist Posture Identification. *ACS Appl. Electron. Mater.* **6**, 9071–9081 (2024).
12. Zhu, Y., et al. Ultrastretchable freezing-tolerant organohydrogels for health monitoring and triboelectric nanogenerator-based gesture recognition. *Nano Energy* **131**, 110261 (2024).
13. Li, C., et al. Chip-less, self-powered electronic patch for wireless sensing and communication. *Nat. Sens.* **1**, 270–278 (2026).
14. Dai, S., et al. A Water-Driven and Low-Damping Triboelectric Nanogenerator Based on Agricultural Debris for Smart Agriculture. *Small* **18**, 2204949 (2022).
15. Zhu, X., et al. Hybrid High-Entropy Vibration Energy Harvester for Self-Powered Sensing in Smart Railways. *Adv. Funt. Mater.* **36**, e76336 (2026).
16. Zhang, C., et al. Magnetic suspension damped hybrid nanogenerator for water wave energy harvesting. *Adv. Energy Mater.* **15**, 2500130 (2025).
17. Su, J. et al. Square-Structured Nylon/Chitosan Oligosaccharide-Based Triboelectric Nanogenerator for Omnidirectional Low-Frequency Energy Harvesting. *ACS Appl. Mater. Interfaces* **17**, 53437–53447 (2025).
18. Qin, S., et al. Ionic and adhesive triboelectric elastomer for normalizing charge polarity and density of contact electrification with general material. *Sci. Adv.* **12**, eaec1580 (2026).
19. Hao, Y. et al. Evaluation and Rehabilitation System for Ulnar-Innervated Muscles Facilitated by Rare Earth Oxide-Enhanced Triboelectric Sensor. *Adv. Mater.* **37**, 2505613 (2025).
20. Zhu, X. et al. A review of charge excitation triboelectric nanogenerator performance enhancement and related applications. *Appl. Mater. Today* **41**, 102492 (2024).
21. Luo, Q., et al. Topological isomerization unlocks exceptional elasticity and strength of cellulosic triboelectric aerogels. *Nat. Commun.* **17**, 5185 (2026).
22. Hao, Y. et al. "Running Toward a Greener Future": Lysine-Enhanced Triboelectric Nanogenerator for Efficient Foot-Motion Energy Harvesting. *Adv. Funct. Mater.* **36**, e19250 (2025).
23. Zou, H., et al. Quantifying the triboelectric series. *Nat. Commun.* **10**, 1427 (2019).
24. Zhao, Z., et al. Selection rules of triboelectric materials for direct-current triboelectric nanogenerator. *Nat. Commun.* **12**, 4686 (2021).
25. Ball, P., et al. Squeezing out power. *Nat. Mater.* **23**, 28 (2024).
26. Yu, Y., et al. Material's selection rules for high performance triboelectric nanogenerators. *Mater. Today* **64**, 61–71 (2023).
27. Meng, H. et al. Triboelectric performances of biodegradable polymers. *Matter* **6**, 4274–4290 (2023).
28. Li, Y., et al. Visualization and standardized quantification of surface charge density for triboelectric materials. *Nat. Commun.* **15**, 6004 (2024).
29. Liu, D., et al. Standardized measurement of dielectric materials' intrinsic triboelectric charge density through the suppression of air breakdown. *Nat. Commun.* **13**, 6019 (2022).

30.Zhang, N., et al. Electronic-structure descriptor-guided design of high-entropy perovskites for efficient CO_2 electrolysis. *Energy Environ. Sci.* **19**, 4222–4232 (2026).
31. Zhang, J., et al. Decoupling Intrinsic Molecular Efficacy From Platform Effects: An Interpretable Machine Learning Framework for Unbiased Perovskite Passivator Discovery. *Adv. Sci.* **13**, e75486 (2026).
32. Li, Y., et al. Machine Learning−Accelerated Discovery of Proton-Conducting 2D Materials for Proton Exchange Membranes. *ACS Nano* **20**, 719–730 (2026).
33.Xiong, X., et al. Development of interfacial oxygen micro-nanobubble technology for hypoxia remediation: Machine-learning assisted optimization and mechanistic insights into sustained oxygen release. *Water Res.* **291**, 125227 (2026).
34.Liu, Z., et al. Universal Catalyst Design Framework for Electrochemical Hydrogen Peroxide Synthesis Facilitated by Local Atomic Environment Descriptors. *Angew. Chem. Int. Ed.* **65**, e18027 (2026).
35. Zhang, Z., et al. Fusion Deep Learning for Predicting Conductivity in Electron-Doped Organic Polymers. *J. Am. Chem. Soc.* **147**, 36233–36243 (2025).

Acknowledgements

This work was supported by the Beijing Key Laboratory of High-Entropy Energy materials and Devices, Beijing Institute of Nanoenergy and Nanosystems (No. GS2025ZD002), Open Project of the State Key Laboratory of Advanced Rail Autonomous Operation at Beijing Jiaotong University (RAO2025ZT004), Beijing Municipal Natural Science Foundation (4232074), National Natural Science Foundation of China (60706031 and 61574015), China Scholarship Council (202507090044).

Author Information

Authors and Affiliations:
a. School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, 100044, P. R. China: Yijun Hao, Jiyu Su, Xiaopeng Zhu, Yunlei Qie, Dong Yang, Wei Su, Hongke Zhang and Xiuhan Li
b. Beijing Institute of Nanoenergy and Nanosystems, Chinese Academy of Sciences, Beijing 101400, P. R. China: Zhong Lin Wang
c. Department of Electrical and Computer Engineering, National University of Singapore, 117576, Singapore: Yijun Hao, Jianlong Hong, Xinliang Zhou Daye Li and Chengkuo Lee
d. Laser Micro/Nano-Fabrication Laboratory, School of Mechanical Engineering, Beijing Institute of Technology, Beijing, 100081, China: Yongjiu Yuan
e. Center for Intelligent Sensors and MEMS (CISM), National University of Singapore, 117583. Singapore: Yijun Hao, Jianlong Hong, Xinliang Zhou and Chengkuo Lee
f. State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing, 100044, P.R. China: Yong Qin and Xiuhan Li
g. State Key Laboratory of Advanced Fiber Materials, College of Materials Science and Engineering, Donghua University, Shanghai, 201620, China: Xinliang Zhou
h. School of Automation and Electrical Engineering, University of Science and Technology, Beijing, 100083, China: Daye Li

Contributions:

Y.H. was responsible for conceptualization, methodology, validation and investigation. Y.H., J.S., X.Zhu, and Y.Q. contributed to the original draft writing, resource acquisition and visualization.

J.S., X.Zhu., and J.H. contributed to software. X.Zhou, D.L., Y.Q. and D.Y. were involved in formal analysis and data curation. W.S., H.Z., C.L., Y.Y., Z.L.W. and X.L. provided review & editing, project administration, supervision and funding acquisition.

Corresponding authors:
Chengkuo Lee, Zhong Lin Wang, Xiuhan Li

Declaration of interests

The authors declare no competing interests.

Journal Pre-proof

Figures

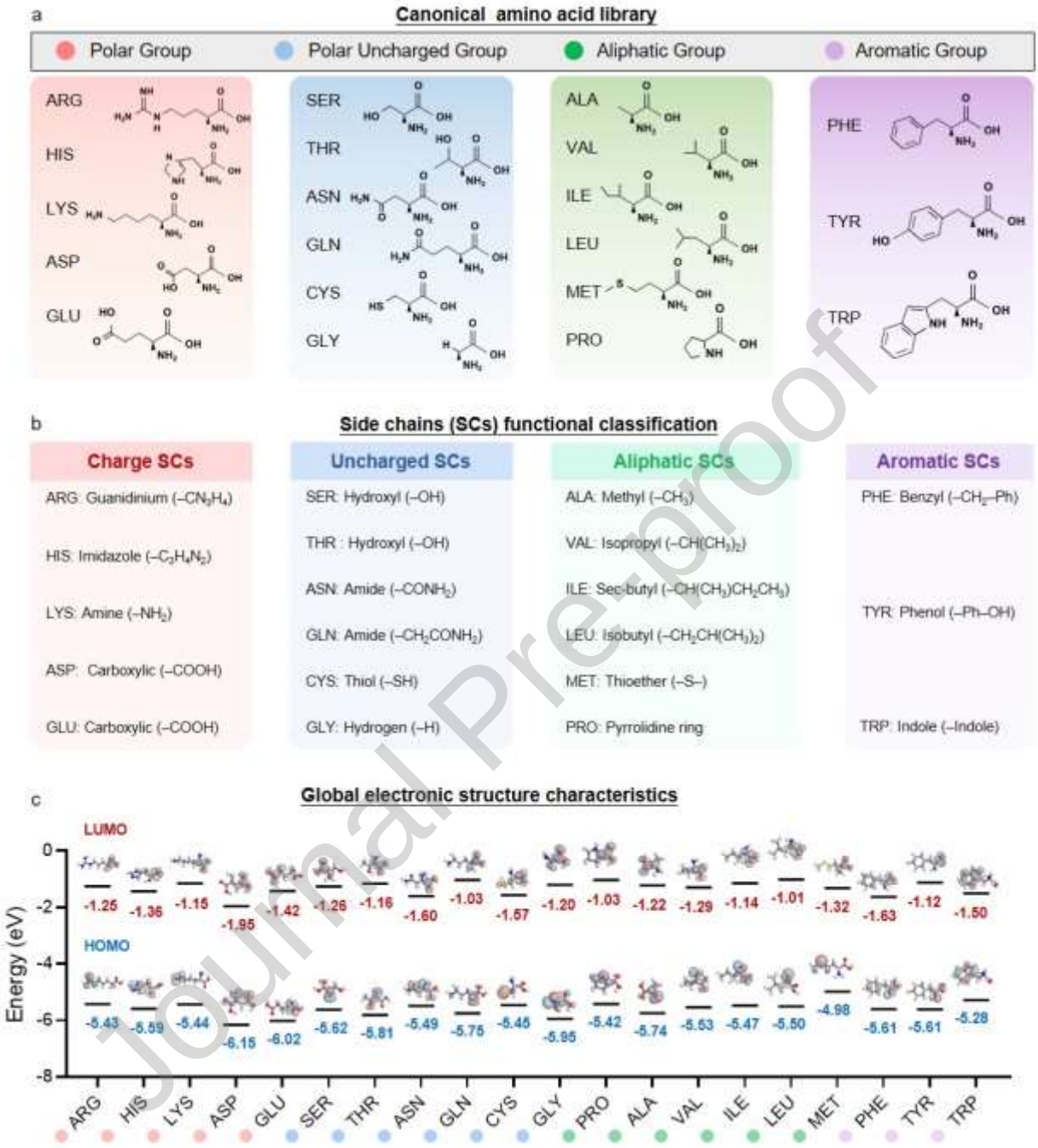


Fig. 1. A molecular library of canonical amino acids. (a) Molecular library comprising all twenty canonical amino acids, classified into four representative groups. (b) Functional 11/12/2026classification of amino acid side chains. (c) Schematic illustration of the subtle electronic variations introduced by different side chains.

Journal Pre-proof

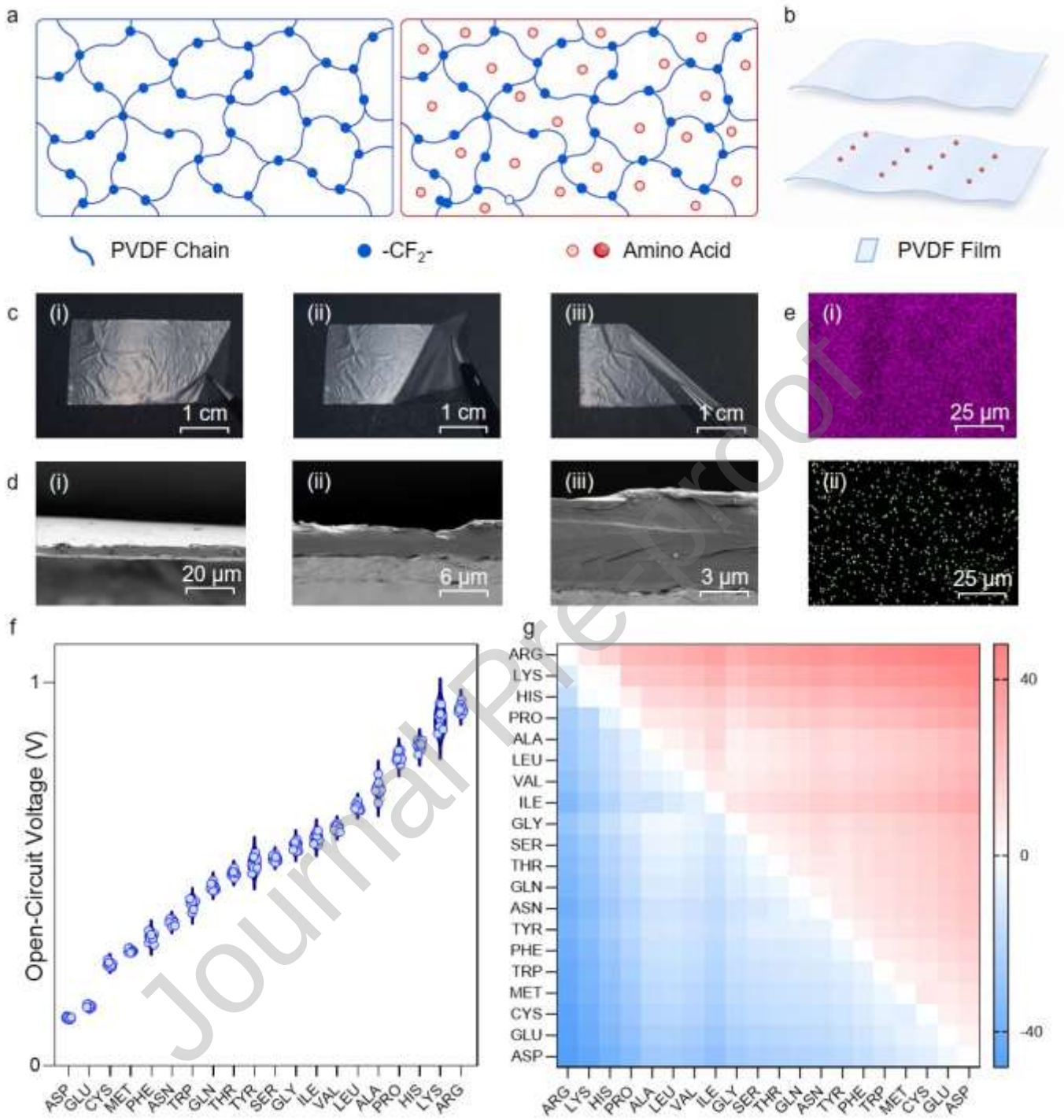


Fig. 2. Establishing the molecular landscape of contact-electrification. (a) Schematic illustration of the AF film. (b) Device configuration using AF and PF as the triboelectric pair. (c) Optical images of the free-standing AF. (d) Cross-sectional SEM images of AF. (e) Elemental mapping of fluorine and nitrogen on the AF surface. (f) Molecular landscape of CE for canonical amino acids. (g) Pairwise CE measurements between representative AF films.

Journal Pre-proof

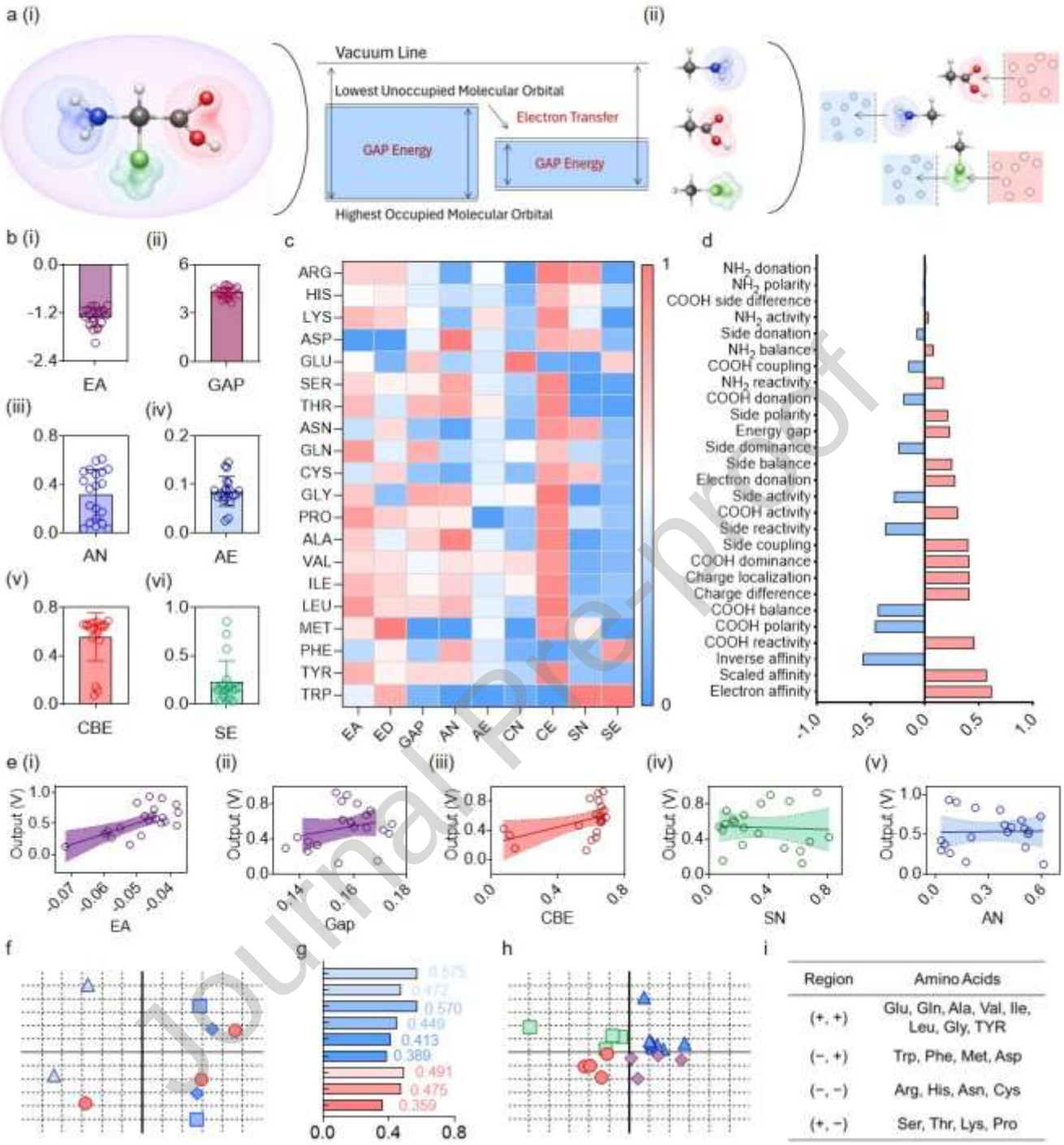


Fig. 3. Descriptor representation of molecular characteristics. (a) Global electronic and site-specific reactive descriptor representations. (b) Distribution of molecular descriptor values across the canonical amino acid library. (c) Normalized molecular descriptor matrix. (d) Semantic interpretation of molecular descriptors. (e) Pearson correlations between representative descriptors (EA, GAP: Hartree) and CE capability. (f) Two-dimensional descriptor embedding. (g) Relative descriptor distances in the embedding space. (h) Molecular embedding of the twenty canonical amino acids. (i) Descriptor-region assignment based on the embedding coordinates.

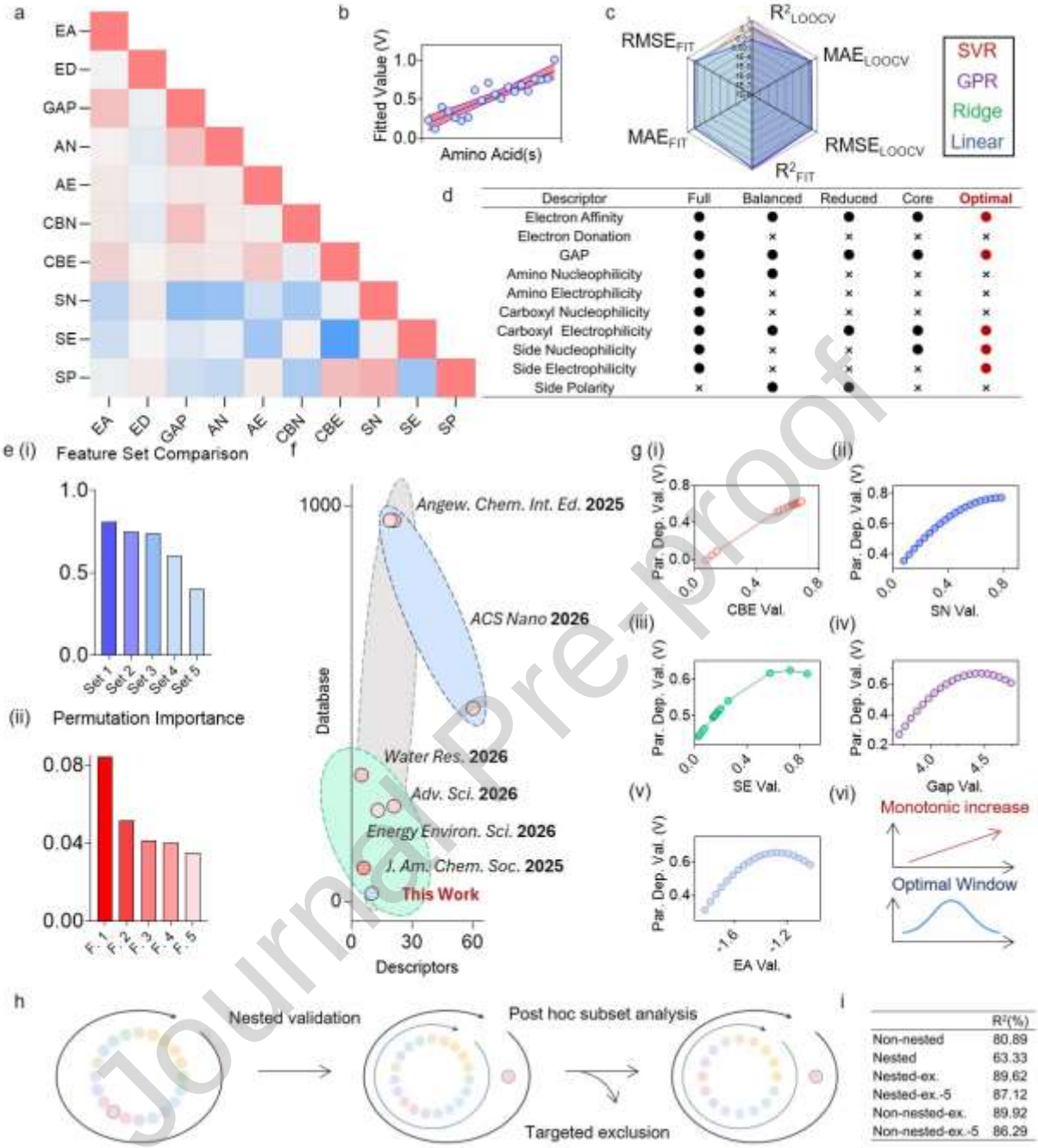


Fig. 4. Learning an interpretable molecular framework. (a) Descriptor coupling matrix. (b) Reconstruction of the CE landscape. (c) Performance comparison of regression models. (d) Hierarchical descriptor representations. (e) Performance contribution of (e i) descriptor-set and (e ii) contribution analysis. (f) Comparison of descriptor dimensionality and dataset size among studies. (g) Descriptor–response relationships: (g i) CBE. (g ii) SN. (g iii) SE. (g iv) GAP. (g v) EA. (g vi) Summary of descriptor–response patterns. h. Validation workflow. i. Comparison across validation and exclusion settings.

Journal Pre-proof

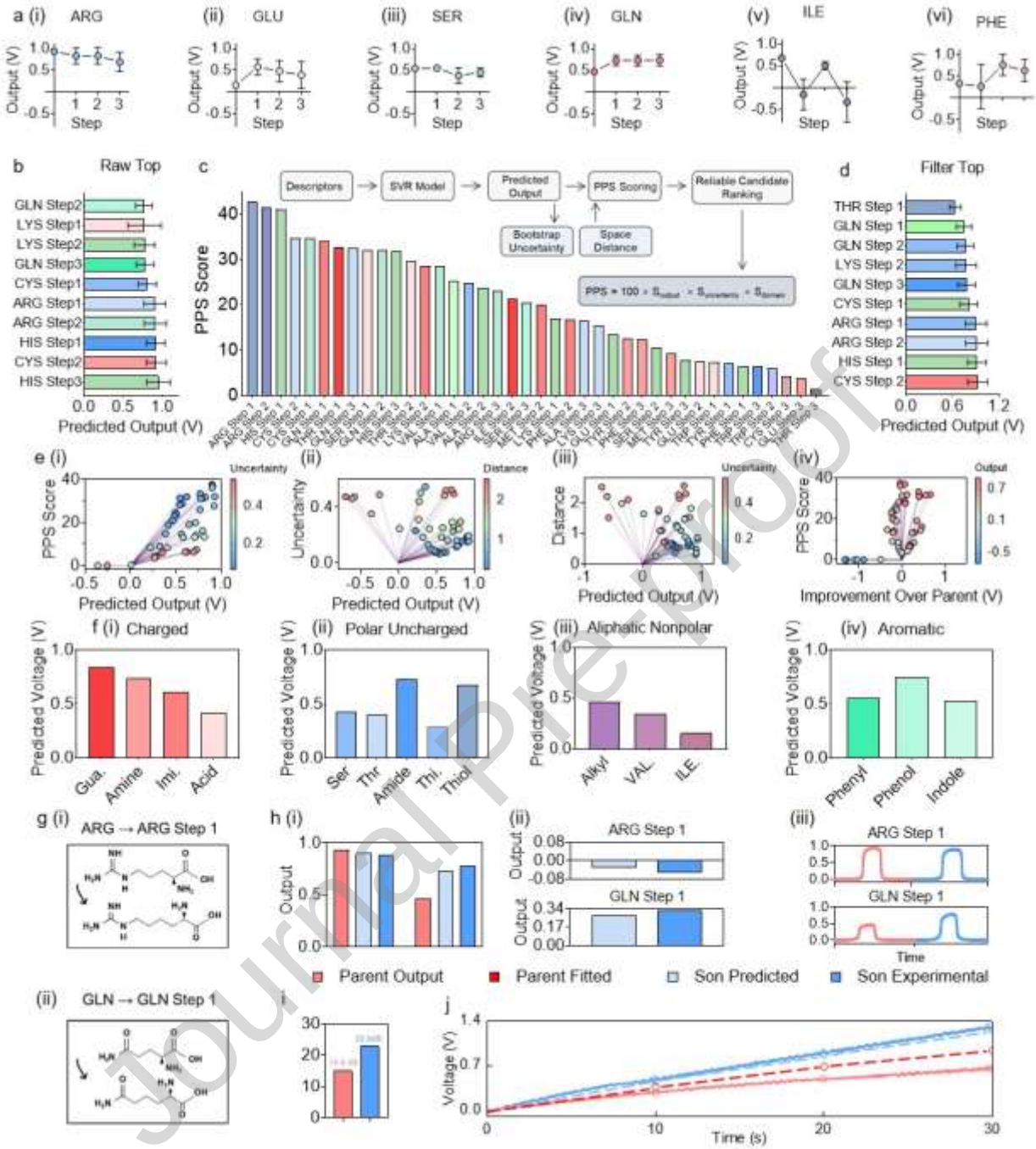


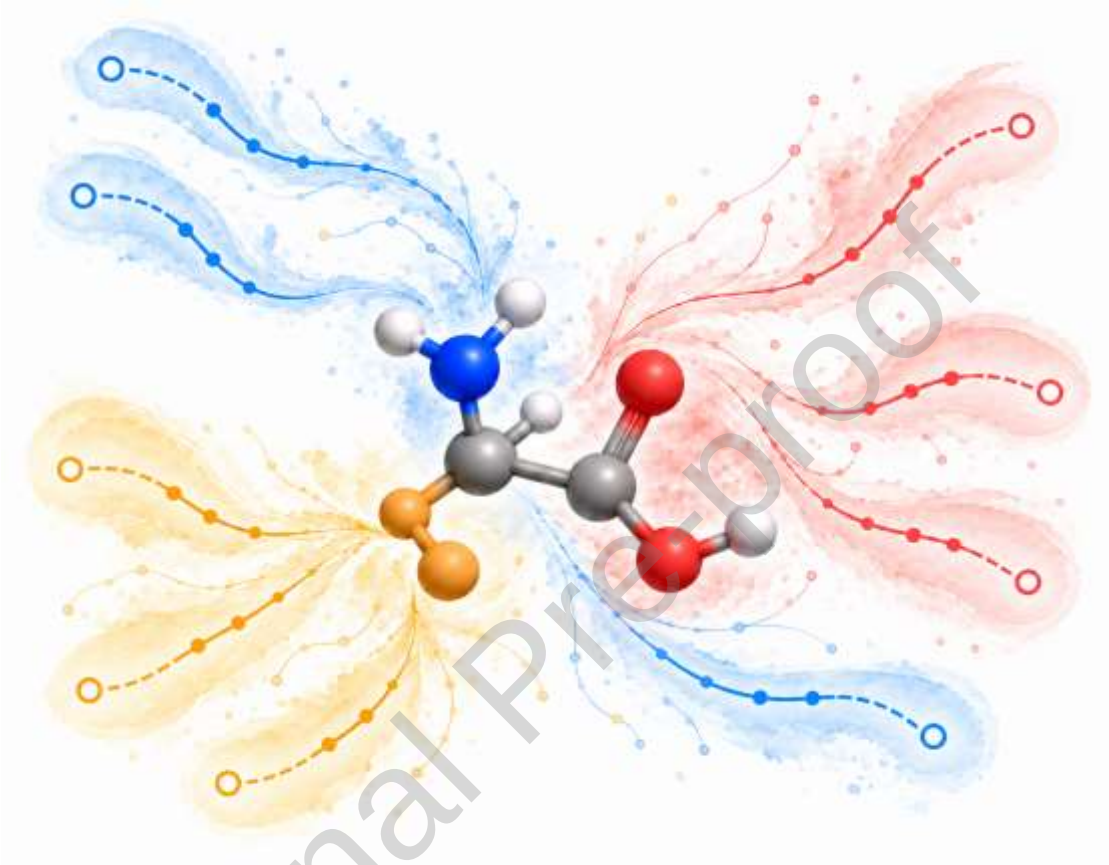
Fig. 5. Framework-guided molecular discovery. (a) Framework-guided prediction of CE capability for (a i) ARG, (a ii) GLU, (a iii) SER, (a iv) GLN, (a v) ILE, (a vi) PHE. (b) Raw Top candidate rank. (c) Illustration of Prediction Priority Score. (d) Filter Top candidate library. (e) Multidimensional relationships among prediction process. (f) Side chain-family-dependent evolutionary trends for (f i) polar, (f ii) nonpolar, (f iii) aliphatic and (f iv) aromatic derivatives. (g) Framework-guided selection of (g i) ARG Step1 and (g ii) GLN Step1 for experimental validation. (h) Experimental validation of framework-guided molecular discovery. (h i) Comparison among parent amino acids, (h ii) changes relative to the parent amino acids, (h iii) comparison between parent and son. (i) Corresponding SNR values. (j) Capacitor-charging performance behaviors.

Journal Pre-proof

Declaration of interests

The authors declare no competing interests.

Graphical abstract



Highlights

- Canonical amino acids serve as a model family for systematic comparison CE capability.
- A quantitative framework is proposed for CE capability of amino acids.
- Molecular structure is linked to the CE capability of amino acids.
- The framework provides a basis for rational selection of triboelectric materials.

Journal Pre-proof