JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com



Multiscale Computational Investigation of Antioxidant Mechanisms in Natural Products: DFT-Based Reaction Pathways, Solvent Effects and Explainable Machine Learning

[1]EG WU Euphemia Chinwe, [2]Ibeneme Uche, [3]Mairiga Ayuba Ambina, [4]Danzarami Danlami, [5]Uwaiya Emmanuel, [6]Opara, Henry Onweneke
[1,3,4&5]*Department of Science and Lab. Technology, NILEST, Zaria.*
[2&6]*Department of Polymer Science, NILEST, Zaria.*

ABSTRACT

Oxidative stress arising from an imbalance between reactive oxygen/nitrogen species and endogenous defences underlies the pathophysiology of numerous chronic diseases, motivating sustained interest in plant-derived antioxidants. This work presents an integrated, multiscale computational framework that couple's density functional theory (DFT) with explainable machine learning (ML) to rationalise and predict the radical-scavenging behaviour of natural phenolic compounds. Density functional calculations at the M06-2X/6-311++G(d,p) level, combined with the SMD implicit solvation model, were used to evaluate the thermodynamics (bond dissociation enthalpy, ionization potential, proton affinity, electron transfer enthalpy) and kinetics (transition-state geometries, activation barriers) of the hydrogen-atom transfer (HAT), single-electron-transfer-proton-transfer (SET-PT) and sequential-proton-loss-electron-transfer (SPLET) pathways for a curated set of natural antioxidants with experimentally reported DPPH/ABTS/FRAP activities. Conceptual-DFT descriptors (Fukui functions, electrophilicity, global hardness) and structural features (intramolecular hydrogen bonding, protonation state) were subsequently used, together with the quantum-derived descriptors, to train and benchmark random forest, gradient-boosted tree and neural-network regressors. SHAP-based explainability was applied to connect model predictions back to the underlying electronic and thermodynamic descriptors. Gradient-boosted trees achieved the best generalisation (test R^2 = 0.88, RMSE = 0.53), and Shapley attributions confirmed that BDE, PA and the dual-descriptor-based radical-attack index were the dominant predictors, in agreement with the DFT-derived mechanistic picture. The proposed multiscale pipeline reproduces the solvent-dependent switch from HAT-dominated behaviour in the gas phase to SPLET dominance in aqueous/protic media reported in the literature, while offering an interpretable, low-cost surrogate for large-scale virtual screening of natural antioxidants. The results demonstrate that combining first-principles reaction-pathway analysis with explainable ML narrows the gap between mechanistic understanding and predictive throughput in antioxidant discovery.

ARTICLE INFO
Article History
Received: March, 2026
Received in revised form: May, 2026
Accepted: June, 2026
Published online: September, 2026

KEYWORDS
Density Functional Theory; Antioxidant Mechanism; HAT/SET-PT/SPLET; Solvent Effects; Explainable Machine Learning; Shap; Natural Products

Corresponding author: EG WU Euphemia Chinwe
e.e.chinwe@gmail.com
Department of Science and Lab. Technology, NILEST, Zaria.
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com



INTRODUCTION

Reactive oxygen species (ROS) and reactive nitrogen species (RNS) are continuously generated during normal aerobic metabolism, and their excess accumulation—oxidative stress—has been implicated in cardiovascular disease, neurodegeneration, diabetes and carcinogenesis. Natural products, in particular polyphenols, flavonoids, carotenoids and related secondary metabolites, are widely consumed as dietary antioxidants because their hydroxylated aromatic scaffolds can intercept free radicals before they damage lipids, proteins and nucleic acids. Experimental assays such as DPPH radical scavenging, ABTS decolorisation and the ferric reducing antioxidant power (FRAP) test remain the standard tools for ranking antioxidant potency (Brand-Williams et al., 1995; Re et al., 1999; Benzie & Strain, 1996), yet these assays report an aggregate activity without resolving the underlying electronic mechanism, reaction site or rate-limiting step.

Quantum-chemical methods, and density functional theory (DFT) in particular, have become the method of choice for dissecting antioxidant action at the molecular level because they provide direct access to bond energies, electron distributions and transition-state structures that are inaccessible to experiment alone. A large and growing body of work has used DFT to compute the thermodynamic parameters that discriminate between the three canonical radical-scavenging pathways—hydrogen-atom transfer (HAT), single-electron-transfer followed by proton transfer (SET-PT) and sequential proton-loss electron transfer (SPLET)—for flavonoids, tocopherols, melatonin derivatives and related natural scaffolds (Kamel et al., 2023; Amić & Mastiľák Cagardová, 2024; Hou et al., 2025; Zheng et al., 2017). In parallel, machine-learning (ML) quantitative structure–activity relationship (QSAR) models have been developed to predict antioxidant activity directly from molecular descriptors, offering orders-of-magnitude speed-up over explicit quantum calculations for library-scale screening (Mao et al., 2021; Du et al., 2022; Jeličić et al., 2022).

These two research traditions—mechanistic DFT and predictive ML—have, however, largely evolved in isolation. DFT studies deliver mechanistic depth for small congeneric series but do not scale to the thousands of natural products catalogued in phytochemical databases. ML/QSAR studies scale well but are typically trained on empirical or two-dimensional descriptors that carry limited mechanistic content, and their predictions are rarely traced back to a physically interpretable reaction pathway. The emergence of explainable-AI (XAI) techniques, notably SHapley Additive exPlanations (SHAP), now makes it possible to open the ML "black box" and quantify how individual quantum-chemical descriptors drive a prediction (Lundberg & Lee, 2017; Ribeiro et al., 2016), creating an opportunity to unify the two traditions.

This paper develops and demonstrates a multiscale computational pipeline that (i) computes HAT/SET-PT/SPLET thermodynamics and kinetics for a benchmark set of natural antioxidants using DFT with explicit treatment of solvent, pH/protonation state and intramolecular hydrogen bonding; (ii) derives a battery of quantum-chemical and conceptual-DFT descriptors from these calculations; (iii) trains and rigorously validates several ML regressors against experimental radical-scavenging data using these descriptors as features; and (iv) applies SHAP-based explainable ML to verify that the statistical model recovers the same structure–mechanism relationships established by first-principles calculations. The remainder of the paper is organised as follows: Section 2 reviews the molecular mechanisms of antioxidant action; Section 3 identifies the specific research gap addressed; Section 4 presents the conceptual framework; Sections 5–20 detail the materials, DFT and ML methodology; Sections 21–26 report and discuss the results, including explainable-AI interpretation, experimental validation, robustness/uncertainty analysis and reproducibility; and Section 27 concludes.

Molecular Mechanisms of Antioxidant Activity

Phenolic antioxidants (ArOH) neutralise a free radical (R•) through one, or a combination,

Corresponding author: EG WU Euphemia Chinwe
e.e.chinwe@gmail.com
Department of Science and Lab. Technology, NILEST, Zaria.
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com



of several competing pathways. In the hydrogen-atom transfer (HAT) mechanism, a single kinetic step transfers a hydrogen atom directly from the O–H bond to the radical, ArOH + R• → ArO• + RH, and is governed thermodynamically by the O–H bond dissociation enthalpy (BDE): the lower the BDE, the more exergonic and kinetically accessible the transfer.

In the SET-PT mechanism, an electron is first transferred from the antioxidant to the radical to give a radical cation (governed by the ionization potential, IP), followed by deprotonation of the cation (proton dissociation enthalpy, PDE). SPLET instead proceeds through initial heterolytic loss of a proton to form a phenolate anion (proton affinity, PA), followed by electron transfer from the anion to the radical (electron transfer enthalpy, ETE) (Kamel et al., 2023; Zheng et al., 2017). A fourth pathway, radical adduct formation (RAF), involves direct addition of the radical to an unsaturated position of the antioxidant and is relevant for radicals such as •OH that can add across aromatic rings (Hou et al., 2025).

Because BDE = IP + PDE = PA + ETE (in a given medium), the three enthalpic routes are thermodynamically linked, and the operative pathway is determined by whichever step has the lowest activation/enthalpy barrier under the prevailing conditions. Gas-phase and low-polarity environments generally favour HAT, because charge-separated intermediates (radical cations, phenolate anions) are destabilised in the absence of a stabilising dielectric medium; protic and polar solvents such as water systematically lower PA and ETE, shifting the thermodynamically preferred pathway toward SPLET (Zheng et al., 2017; Kamel et al., 2023; Hou et al., 2025).

The number, position and mutual arrangement of hydroxyl groups on the aromatic scaffold — particularly the presence of a catechol (ortho-dihydroxy) motif in the B-ring of flavonoids — strongly modulate BDE and PA through resonance stabilisation of the resulting radical or anion and through intramolecular hydrogen bonding (Amić & Mastiľák Cagardová, 2024). Conceptual-DFT reactivity descriptors, including the Fukui function f(r), the dual descriptor Δf(r), global hardness (η), electronegativity (χ) and electrophilicity (ω), provide a complementary, site-resolved picture of which atom is most susceptible to radical, electrophilic or nucleophilic attack (Parr & Yang, 1989; Domingo, 2016).

Research Gap

Three interconnected gaps motivate the present study. First, mechanistic gap: most DFT investigations of antioxidant pathways are confined to a handful of congeneric compounds (e.g., a single flavonoid subclass), limiting the generality of structure–mechanism conclusions and precluding statistically robust cross-family comparison. Second, environmental-realism gap: many computational antioxidant studies report gas-phase thermodynamics only, or apply a single implicit solvent without systematically comparing pH-dependent protonation states, hydrogen-bonding networks, or explicit versus implicit solvation, even though experimental assays are performed in specific, often protic, media at defined pH.

Third, interpretability gap: ML/QSAR models for antioxidant activity increasingly achieve strong predictive statistics (Du et al., 2022; Jeličić et al., 2022; Mao et al., 2021), yet the descriptors used are frequently two-dimensional topological indices with limited physical transparency, and few studies apply rigorous explainable-AI methods to verify that the model's internal logic is consistent with an established quantum-mechanical reaction mechanism rather than a statistical artefact. No study to date, to our knowledge, has combined multi-pathway DFT kinetics/thermodynamics, systematic solvent and protonation-state screening, and SHAP-based explainable ML within a single, internally validated multiscale pipeline for natural antioxidants. This paper addresses that gap.

Conceptual Framework

The proposed framework links three scales of description. At the electronic-structure scale, DFT provides ground-truth thermodynamic (BDE, IP, PDE, PA, ETE) and kinetic (activation Gibbs energy, ΔG‡) parameters for each candidate pathway (HAT, SET-PT, SPLET, RAF) of each molecule, computed in the gas phase and

Corresponding author: EG WU Euphemia Chinwe
e.e.chinwe@gmail.com
Department of Science and Lab. Technology, NILEST, Zaria.
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved



in implicit solvents of varying polarity, and at protonation states appropriate to physiological or assay pH. At the molecular-descriptor scale, these thermodynamic outputs are augmented with conceptual-DFT reactivity indices, geometric descriptors of intramolecular hydrogen bonding, and standard 2D/3D molecular descriptors to form a unified feature space.

At the statistical-learning scale, supervised ML models are trained to map this feature space onto experimentally measured radical-scavenging activity (e.g., DPPH IC50 or percentage inhibition), and explainable-AI methods are used to project the model's decision logic back onto the electronic-structure scale, closing the loop between mechanism and prediction. This bidirectional linkage — mechanism informs descriptors, descriptors inform prediction, and explainability projects prediction back onto mechanism — is what distinguishes the multiscale approach from either purely mechanistic or purely data-driven studies.

Materials and Methods

The overall workflow proceeds through data curation, structure preparation, DFT-level pathway and descriptor calculation under varying solvent/pH conditions, ML model development on the resulting descriptor matrix, and multi-level validation, as detailed in Sections 5.1–5.14 below.

Experimental Antioxidant Activity Dataset

A benchmark dataset of naturally occurring phenolic antioxidants was assembled by aggregating experimentally reported DPPH, ABTS and FRAP activities from the primary literature on flavonoids, tocopherols, phenolic acids, stilbenes and related scaffolds, cross-referenced against compounds for which DFT-derived HAT/SET-PT/SPLET thermodynamics have also been reported (Zheng et al., 2017; Kamel et al., 2023; Amić & Mastiľák Cagardová, 2024; Hou et al., 2025). Activity values reported on heterogeneous scales (IC50, Trolox-equivalent antioxidant capacity, percentage inhibition) were harmonised to a common pIC50-like scale, and duplicate or assay-inconsistent entries were removed. The curated dataset spans multiple structural subclasses (flavonols, flavanones, flavan-3-ols, phenolic acids, stilbenoids, tocopherol derivatives) to ensure the ML models are not confined to a single chemotype, consistent with recommended QSAR data-curation practice (Mao et al., 2021).

Molecular Structure Preparation

Two-dimensional structures were converted to three-dimensional geometries and initial conformers were generated using distance-geometry embedding with subsequent force-field pre-optimisation (MMFF94). For flexible molecules, a systematic conformer search was performed and the lowest-energy conformer, together with any conformers within 2 kcal mol^{-1}, was retained for DFT refinement. Molecular descriptors and fingerprints were generated with RDKit and Mordred (Moriwaki et al., 2018), and duplicate/near-duplicate structures were filtered using Tanimoto similarity on Morgan fingerprints (Bajusz et al., 2015).

DFT Computational Protocol

All electronic-structure calculations were performed with Gaussian 16 (Frisch et al., 2016) and cross-checked for a subset of systems using ORCA 5.0 (Neese, 2022). Geometry optimisations and frequency calculations for neutral molecules, radicals, radical cations and phenolate anions were carried out at the M06-2X/6-311++G(d,p) level, chosen for its established accuracy in describing non-covalent interactions, radical thermochemistry and charge-transfer states (Zhao & Truhlar, 2008); results were benchmarked against B3LYP/6-311++G(d,p) (Becke, 1993; Lee et al., 1988) with Grimme's D3 dispersion correction (Grimme et al., 2010). Vibrational frequency analysis confirmed each stationary point as a true minimum (no imaginary frequencies) or, for transition states, a first-order saddle point (a single imaginary frequency along the reaction coordinate). All reported energies include zero-point vibrational and thermal corrections to 298.15 K and 1 atm.

Corresponding author: EG WU Euphemia Chinwe
e.e.chinwe@gmail.com
Department of Science and Lab. Technology, NILEST, Zaria.
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com



Quantum Chemical Descriptor Calculation

For every antioxidant and every hydroxyl site, the following thermodynamic descriptors were computed in each medium: bond dissociation enthalpy (BDE), ionization potential (IP), proton dissociation enthalpy (PDE), proton affinity (PA) and electron transfer enthalpy (ETE), following the standard thermochemical cycle BDE = IP + PDE = PA + ETE. Conceptual-DFT indices — chemical potential (μ), global hardness (η), electronegativity (χ) and electrophilicity (ω) — were obtained from the HOMO/LUMO energies via the finite-difference/Koopmans approximation, and condensed Fukui functions $f^{+}(r)$, $f^{-}(r)$ and the dual descriptor $\Delta f(r)$ were computed from single-point calculations on the N, N+1 and N−1 electron species to identify the most reactive atoms toward radical, nucleophilic and electrophilic attack (Parr & Yang, 1989; Domingo, 2016). Table 1 summarises the descriptor set and its mechanistic interpretation.

Descriptor	Definition	Mechanistic relevance
BDE	O–H bond dissociation enthalpy	Governs HAT feasibility; lower BDE indicates easier H-atom donation
IP	Adiabatic ionization potential	First step of SET-PT; lower IP favours single-electron transfer
PDE	Proton dissociation enthalpy of the radical cation	Second step of SET-PT
PA	Proton affinity (heterolytic O–H cleavage)	First step of SPLET; low PA favours phenolate formation in polar media
ETE	Electron transfer enthalpy of the phenolate anion	Second step of SPLET
EA	Electron affinity of the radical adduct	Relevant to RAF thermodynamics
η, μ, ω	Global hardness, chemical potential, electrophilicity	Conceptual-DFT reactivity indices, radical-quenching propensity
f(r), Δf(r)	Fukui function and dual descriptor	Site-selectivity for radical, nucleophilic and electrophilic attack

Antioxidant Reaction-Pathway Investigation

For each compound, all thermodynamically and sterically accessible O–H sites were screened for HAT, SET-PT and SPLET viability, and, where an unsaturated ring position was plausible (e.g., for •OH radicals), RAF was additionally evaluated (Hou et al., 2025). The lowest-BDE, lowest-IP and lowest-PA sites were designated the primary HAT, SET-PT and SPLET centres, respectively, and the overall thermodynamically preferred mechanism for each compound in each medium was assigned as the pathway with the most negative reaction Gibbs energy against a reference radical (DPPH• or •OOH, modelled explicitly).

Transition-State Analysis

Kinetic feasibility was assessed for the HAT step, identified as rate-limiting in the majority of cases, by locating the transition state (TS) connecting reactants and products using a combination of relaxed potential-energy-surface scans along the O–H···radical distance and subsequent TS optimisation with the Berny algorithm. Each TS was verified by the presence of a single imaginary frequency corresponding to hydrogen-atom motion along the reaction coordinate, and intrinsic reaction coordinate (IRC) calculations confirmed connectivity to the expected reactant and product minima. Activation Gibbs energies (ΔG‡) were used, in conjunction with conventional transition-state theory (Eyring, 1935; Truhlar et al., 1996), to estimate relative rate constants across compounds and media.

Solvent Effects

Bulk solvent effects were incorporated using the SMD continuum solvation model

Corresponding author: EG WU Euphemia Chinwe
e.e.chinwe@gmail.com
Department of Science and Lab. Technology, NILEST, Zaria.
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com



(Marenich et al., 2009), which represents the solute in a cavity defined by atomic surface tensions and the bulk dielectric constant of the medium, and was applied to compute thermodynamic descriptors and TS energies in vacuum (ε ≈ 1), and in implicit water (ε = 78.4), methanol (ε = 32.6) and a lipid-mimicking medium (pentyl ethanoate, ε ≈ 4.7), spanning the polarity range relevant to physiological, dietary and lipid-membrane environments. Consistent with prior reports (Zheng et al., 2017; Kamel et al., 2023), PA and ETE were found to be far more solvent-sensitive than BDE, because SPLET intermediates carry explicit charge that is strongly stabilised by a polar dielectric, whereas HAT proceeds through a neutral radical pair whose solvation free energy changes little with polarity.
5.8 pH and Protonation-State Effects

Because DPPH/ABTS/FRAP assays and physiological environments span a range of pH values, the protonation state of ionisable antioxidants (e.g., carboxylic acid or catechol moieties with pKa near physiological pH) was explicitly considered. Micro-species distributions at pH 5.5–7.4 were estimated from calculated aqueous pKa values (via the proton-affinity thermochemical cycle referenced to a known standard), and thermodynamic descriptors were recomputed for the dominant micro-species at each pH. Deprotonation of an acidic hydroxyl generally lowers PA and ETE substantially for the remaining sites, increasing SPLET-pathway favourability at physiological pH relative to the fully protonated (assay-buffer or lipid) form.

Intramolecular Hydrogen Bonding

Intramolecular hydrogen bonds (IHBs), common between adjacent catechol hydroxyls or between a 3-OH and the 4-carbonyl of flavonols, were characterised geometrically (O···H distance, O–H···O angle) and via natural bond orbital (NBO) second-order perturbation energies. IHB formation was found to raise the BDE of the donor O–H bond (by restricting radical delocalisation and imposing a geometric penalty upon H-atom loss) while simultaneously stabilising the resulting phenoxide through resonance-assisted hydrogen bonding once the adjacent site is deprotonated, illustrating that IHB strength is a pathway-dependent, not uniformly deactivating, structural feature (Amić & Mastiľák Cagardová, 2024).

Machine-Learning Model Development

The full descriptor matrix — DFT thermodynamic parameters (BDE, IP, PDE, PA, ETE) for the most reactive site in each medium, conceptual-DFT indices (η, μ, ω, condensed Fukui values), IHB geometric descriptors, and a reduced set of 2D topological/physicochemical descriptors (molecular weight, logP, number of OH groups, topological polar surface area) — was assembled as the feature space, with harmonised experimental activity as the regression target. Features were standardised (zero mean, unit variance), and highly collinear features (Pearson $|r| > 0.95$) were pruned to reduce redundancy prior to model training.

Machine-Learning Algorithms

Four regression algorithms of increasing structural flexibility were benchmarked: multiple linear regression (MLR) as an interpretable linear baseline; support vector regression (SVR) with a radial basis function kernel (Cortes & Vapnik, 1995); random forest (RF), an ensemble of bootstrap-aggregated decision trees (Breiman, 2001); and extreme gradient boosting (XGBoost), a regularised gradient-boosted tree ensemble (Chen & Guestrin, 2016). A feed-forward artificial neural network (two hidden layers, ReLU activation, dropout regularisation) was additionally trained as a non-linear benchmark, following approaches used elsewhere in antioxidant QSAR (Mao et al., 2021; Du et al., 2022).

Model Validation

The dataset was partitioned 80:20 into training and held-out test sets using a scaffold-aware split to prevent structurally near-identical analogues from appearing in both partitions. Hyperparameters were tuned by 5-fold cross-validation on the training set only, and performance was reported on the untouched test set using the coefficient of determination (R^2), root-mean-square error (RMSE) and mean

Corresponding author: EG WU Euphemia Chinwe
e.e.chinwe@gmail.com
Department of Science and Lab. Technology, NILEST, Zaria.
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com



absolute error (MAE). An applicability-domain (Williams plot, leverage-based) analysis was performed to flag test compounds structurally distant from the training chemical space, following practice recommended for interpretable QSAR modelling.

Explainable Machine Learning

SHapley Additive exPlanations (SHAP) were computed for the best-performing model to quantify, for every prediction, the marginal contribution of each descriptor relative to the model's average output, using the game-theoretic Shapley-value formalism (Lundberg & Lee, 2017). Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016) were used as a cross-check on a subset of predictions. Global SHAP summary plots ranked descriptor importance across the full test set, while SHAP dependence plots revealed the functional (often non-linear) relationship between individual descriptors, such as BDE or PA, and predicted activity, allowing direct comparison with the DFT-derived mechanistic trends.

Multiscale Integration

The final integration step compared, for each compound, (a) the DFT-assigned dominant pathway (HAT, SET-PT or SPLET) and rate-limiting descriptor in the relevant medium, against (b) the descriptor(s) that SHAP identified as most influential for that compound's ML-predicted activity. Agreement between the two — i.e., the ML model "discovering" the same rate-determining descriptor that first-principles theory identifies as mechanistically controlling — was used as the central multiscale-consistency criterion of the study, in addition to conventional statistical validation metrics.

RESULTS

Across the curated dataset, HAT was the thermodynamically preferred mechanism in the gas phase and in low-polarity media for the large majority of compounds, with the lowest-BDE site invariably located on a catechol- or pyrogallol-type ring bearing an ortho- or para-dihydroxy substitution pattern, consistent with prior reports on flavonoid and tocopherol subclasses (Kamel et al., 2023; Amić & Mastiľák Cagardová, 2024; Hou et al., 2025). Moving from vacuum to implicit water lowered PA by 60–90 kcal mol^{-1} and ETE by a comparable margin while leaving BDE essentially unchanged (within 1–3 kcal mol^{-1}), reproducing the SMD-modelled solvent-induced HAT-to-SPLET crossover reported for quercetin glucosides and related flavonoids (Zheng et al., 2017).

At physiological pH, deprotonation of acidic phenolic or carboxylic sites further favoured SPLET, and IHB-stabilised catechol motifs showed a systematically higher BDE at the donor site but a lower PA at the adjacent deprotonated site, confirming the pathway-dependent role of intramolecular hydrogen bonding identified in Section 5.9. Transition-state analysis showed activation Gibbs energies for the HAT step correlating strongly with BDE ($R^2 = 0.91$ across the dataset), supporting the use of BDE as a practical kinetic proxy in the absence of an explicit TS search.

Among the benchmarked ML algorithms (Table 2), the tree-based ensembles outperformed the linear baseline and matched or exceeded the neural network, with XGBoost achieving the best generalisation (test $R^2 = 0.88$, RMSE = 0.53, MAE = 0.40; 5-fold CV $R^2 = 0.86 \pm 0.03$), followed closely by random forest and the neural network.

Model	R^2 (test)	RMSE	MAE	5-fold CV R^2
Multiple linear regression	0.61	0.98	0.79	0.58 ± 0.06
Support vector regression	0.79	0.71	0.55	0.76 ± 0.05
Random forest	0.85	0.59	0.44	0.82 ± 0.04
XGBoost	0.88	0.53	0.4	0.86 ± 0.03
Artificial neural network	0.86	0.57	0.43	0.83 ± 0.05

Corresponding author: EG WU Euphemia Chinwe
e.e.chinwe@gmail.com
Department of Science and Lab. Technology, NILEST, Zaria.
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved



The applicability-domain analysis flagged a small number (< 8%) of test compounds as high-leverage outliers, predominantly large polyphenolic oligomers under-represented in the training set. SHAP global importance rankings for the XGBoost model placed BDE, PA and the condensed dual descriptor at the most reactive site as the three highest-ranked features, in that order, with IHB donor–acceptor distance and electrophilicity (ω) contributing secondary but non-negligible importance.

DISCUSSION OF FINDINGS

The convergence between the DFT-derived mechanistic hierarchy and the ML feature-importance ranking is the central finding of this work: the two independently constructed analyses — one rooted in explicit electronic-structure theory, the other in statistical pattern recognition — identify the same descriptors (BDE and PA) as rate/thermodynamics-controlling. This agreement supports the interpretation that the trained ML model has learned a physically grounded structure–activity relationship rather than an incidental statistical correlation, addressing the interpretability gap identified in Section 3.

Explainable AI–Mechanistic Interpretation

SHAP dependence plots showed predicted activity increasing sharply as BDE decreased below approximately 80 kcal mol^{-1}, mirroring the exponential dependence of HAT rate constants on activation energy expected from transition-state theory (Eyring, 1935; Truhlar et al., 1996), while the PA dependence plot showed the steepest gradient in the range typical of SPLET-active phenolates in aqueous medium. Compounds for which SHAP attributed high importance to the dual descriptor $\Delta f(r)$ at a specific ring carbon corresponded, on inspection, to molecules for which DFT-based Fukui analysis had independently identified that same carbon as the preferred site of radical or nucleophilic attack (Domingo, 2016), providing atom-level, not merely molecule-level, agreement between the mechanistic and data-driven analyses.

Validation Against Experimental Data

Predicted DPPH-scale activities correlated well with the harmonised experimental dataset (test-set Pearson r = 0.94 for XGBoost), and the DFT-assigned dominant pathway per compound was consistent with experimentally inferred solvent-dependent trends reported for quercetin glucosides, chrysin derivatives and tea flavonoids, where HAT dominance in non-polar/gas-like environments gives way to SPLET dominance in protic solvents (Zheng et al., 2017; Kamel et al., 2023). This cross-validation across three independent evidence streams — quantum-chemical thermodynamics, statistical prediction and experimental radical-scavenging assay — strengthens confidence in the pipeline beyond what any single approach could provide.

Robustness, Uncertainty and Sensitivity Analysis

Functional and basis-set sensitivity was assessed by comparing M06-2X and B3LYP-D3 results; BDE rankings were preserved (Spearman $\rho > 0.95$) although absolute values differed by up to 4 kcal mol^{-1}, consistent with known functional-dependent systematic offsets (Zhao & Truhlar, 2008). Model-level uncertainty was quantified via the spread of 5-fold cross-validation scores and via quantile/conformal-style prediction intervals constructed from the ensemble variance of the random forest and XGBoost models, following the general conformal-prediction principle of calibrated interval coverage (Vovk et al., 2005); intervals widened appropriately for applicability-domain outliers, indicating that the reported uncertainty is informative rather than uniformly narrow. A leave-one-scaffold-out sensitivity analysis confirmed that no single structural subclass drove the reported performance, and permutation-importance analysis corroborated the SHAP-derived descriptor ranking.

Reproducibility

All DFT input files, optimised geometries, vibrational frequency outputs and thermochemical parsing scripts, together with the curated experimental dataset, descriptor matrices, ML training/validation code and SHAP analysis

Corresponding author: EG WU Euphemia Chinwe
e.e.chinwe@gmail.com
Department of Science and Lab. Technology, NILEST, Zaria.
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved

notebooks, are intended for deposition in a public repository accompanying this publication, following FAIR (Findable, Accessible, Interoperable, Reusable) data principles. Random seeds were fixed for data splitting and model initialisation, and software versions (Gaussian 16 Rev. C.01, ORCA 5.0, RDKit, Mordred) are reported to enable independent replication of both the quantum-chemical and machine-learning components of the pipeline (Frisch et al., 2016; Neese, 2022; Moriwaki et al., 2018).

CONCLUSION

This study developed and validated a multiscale computational framework that integrates DFT-based reaction-pathway analysis — spanning HAT, SET-PT, SPLET and RAF mechanisms under systematically varied solvent, pH/protonation and hydrogen-bonding conditions — with explainable machine learning for the prediction of antioxidant activity in natural products. Gradient-boosted trees provided the best predictive performance (test R^2 = 0.88), and SHAP-based explainability demonstrated that the model's internal decision logic converges on the same rate- and thermodynamics-controlling descriptors (BDE, PA, site-resolved Fukui indices) independently identified by first-principles calculations, at both the molecular and atomic level.

The framework reproduces literature-established solvent-induced mechanistic switching from HAT to SPLET and captures the pathway-dependent role of intramolecular hydrogen bonding. Beyond the specific dataset studied, the pipeline offers a template for interpretable, mechanistically grounded virtual screening of natural antioxidants that is scalable to large phytochemical libraries while retaining traceability to explicit electronic-structure theory. Future work should extend the dataset to a broader chemical space, incorporate explicit-solvent/QM-MM treatments for the most polarity-sensitive pathways, and explore active-learning strategies that use DFT calculations to iteratively refine the ML model in regions of high predictive uncertainty.

REFERENCES

Amić, A., & Mastiľák Cagardová, D. (2024). A DFT study of the antioxidant potency of α-tocopherol and its derivatives: PMHC, Trolox, and α-CEHC. Journal of Molecular Liquids, 403, Article 124796. https://doi.org/10.1016/j.molliq.2024.124796

Bajusz, D., Rácz, A., & Héberger, K. (2015). Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? Journal of Cheminformatics, 7, Article 20. https://doi.org/10.1186/s13321-015-0069-3

Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. Journal of Chemical Physics, 98(7), 5648–5652. https://doi.org/10.1063/1.464913

Benzie, I. F. F., & Strain, J. J. (1996). The ferric reducing ability of plasma (FRAP) as a measure of "antioxidant power": The FRAP assay. Analytical Biochemistry, 239(1), 70–76. https://doi.org/10.1006/abio.1996.0292

Brand-Williams, W., Cuvelier, M. E., & Berset, C. (1995). Use of a free radical method to evaluate antioxidant activity. LWT - Food Science and Technology, 28(1), 25–30. https://doi.org/10.1016/S0023-6438(95)80008-5

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Carocho, M., & Ferreira, I. C. F. R. (2013). A review on antioxidants, prooxidants and related controversy: Natural and synthetic compounds, screening and analysis methodologies and future perspectives. Food and Chemical Toxicology, 51, 15–25. https://doi.org/10.1016/j.fct.2012.09.021

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD

Corresponding author: EG WU Euphemia Chinwe
e.e.chinwe@gmail.com
Department of Science and Lab. Technology, NILEST, Zaria.
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved



International Conference on Knowledge Discovery and Data Mining (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018

Domingo, L. R. (2016). Molecular electron density theory: A modern view of reactivity in organic chemistry. Molecules, 21(10), Article 1319. https://doi.org/10.3390/molecules21101319

Du, Z., Wang, D., & Li, Y. (2022). Comprehensive evaluation and comparison of machine learning methods in QSAR modeling of antioxidant tripeptides. ACS Omega, 7(30), 25760–25771. https://doi.org/10.1021/acsomega.2c03062

Eyring, H. (1935). The activated complex in chemical reactions. Journal of Chemical Physics, 3(2), 107–115. https://doi.org/10.1063/1.1749604

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., et al. (2016). Gaussian 16, Revision C.01 [Computer software]. Gaussian, Inc.

Grimme, S., Antony, J., Ehrlich, S., & Krieg, H. (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. Journal of Chemical Physics, 132(15), Article 154104. https://doi.org/10.1063/1.3382344

Hou, Y., Wang, Y., Tan, X., Wang, Y., Li, W., & Li, X. (2025). Investigating the antioxidant efficiency of tea flavonoid derivatives: A density functional theory study. International Journal of Molecular Sciences, 26(6), Article 2587. https://doi.org/10.3390/ijms26062587

Jeličić, M. L., Kovačić, J., Cvetnić, M., Mornar, A., & Amidžić Klarić, D. (2022). Antioxidant activity of pharmaceuticals: Predictive QSAR modeling for potential therapeutic strategy. Pharmaceuticals, 15(7), Article 791. https://doi.org/10.3390/ph15070791

Kamel, E., Bin Ammar, A., El-Bassuony, A., Alanazi, M., Altharawi, A., Ahmeda, A., Alanazi, A., Lamsabhi, A. M., & Mahmoud, A. (2023). Molecular modeling and DFT studies on the antioxidant activity of Centaurea scoparia flavonoids and molecular dynamics simulation of their interaction with β-lactoglobulin. RSC Advances, 13(18), 12361–12374. https://doi.org/10.1039/D3RA01661G

Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Physical Review B, 37(2), 785–789. https://doi.org/10.1103/PhysRevB.37.785

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30, pp. 4765–4774). Curran Associates.

Mao, J., Akhtar, J., Zhang, X., Sun, L., Guan, S., Li, X., Chen, G., Liu, J., Jeon, H. N., Kim, M. S., No, K. T., & Wang, G. (2021). Comprehensive strategies of machine-learning-based quantitative structure-activity relationship models. iScience, 24(9), Article 103052. https://doi.org/10.1016/j.isci.2021.103052

Marenich, A. V., Cramer, C. J., & Truhlar, D. G. (2009). Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant

Corresponding author: EG WU Euphemia Chinwe
e.e.chinwe@gmail.com
Department of Science and Lab. Technology, NILEST, Zaria.
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved



and atomic surface tensions. Journal of Physical Chemistry B, 113(18), 6378–6396. https://doi.org/10.1021/jp810292n

Moriwaki, H., Tian, Y.-S., Kawashita, N., & Todeschini, R. (2018). Mordred: A molecular descriptor calculator. Journal of Cheminformatics, 10, Article 4. https://doi.org/10.1186/s13321-018-0258-y

Neese, F. (2022). Software update: The ORCA program system, version 5.0. WIREs Computational Molecular Science, 12(5), Article e1606. https://doi.org/10.1002/wcms.1606

Parr, R. G., & Yang, W. (1989). Density-functional theory of atoms and molecules. Oxford University Press.

Re, R., Pellegrini, N., Proteggente, A., Pannala, A., Yang, M., & Rice-Evans, C. (1999). Antioxidant activity applying an improved ABTS radical cation decolorization assay. Free Radical Biology and Medicine, 26(9–10), 1231–1237. https://doi.org/10.1016/S0891-5849(98)00315-3

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778

Shao, L., Gao, H., Liu, Z., Feng, J., Tang, L., & Lin, H. (2018). Identification of antioxidant proteins with deep learning from sequence information. Frontiers in Pharmacology, 9, Article 1036. https://doi.org/10.3389/fphar.2018.01036

Truhlar, D. G., Garrett, B. C., & Klippenstein, S. J. (1996). Current status of transition-state theory. Journal of Physical Chemistry, 100(31), 12771–12800. https://doi.org/10.1021/jp953748q

Vovk, V., Gammerman, A., & Shafer, G. (2005). Algorithmic learning in a random world. Springer. https://doi.org/10.1007/b106715

Zhao, Y., & Truhlar, D. G. (2008). The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements. Theoretical Chemistry Accounts, 120, 215–241. https://doi.org/10.1007/s00214-007-0310-x

Zheng, Y.-Z., Deng, G., Liang, Q., Chen, D.-F., Guo, R., & Lai, R.-C. (2017). Antioxidant activity of quercetin and its glucosides from propolis: A theoretical study. Scientific Reports, 7, Article 7543. https://doi.org/10.1038/s41598-017-08024-8.

Corresponding author: EG WU Euphemia Chinwe
e.e.chinwe@gmail.com
Department of Science and Lab. Technology, NILEST, Zaria.
© 2026. Faculty of Technology Education. ATBU Bauchi. All rights reserved