

Computational Characterization of Flavonoid Recognition of CAG-Repeat Mismatch Motifs Using Molecular Docking, Global Reactivity Descriptors, and Pharmacokinetic Profiling

Kiara Garcia Homs

New Mexico Highlands University

Bijili Navine Nanjo

New Mexico Highlands University

Nabanita Saikia

nsaikia@nmhu.edu

New Mexico Highlands University

Article

Keywords:

Posted Date: September 2nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10751470/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Computational Characterization of Flavonoid Recognition of CAG-Repeat Mismatch Motifs Using Molecular Docking, Global Reactivity Descriptors, and Pharmacokinetic Profiling

Kiara Garcia Homs, Bijili Navine Nanjo, Nabanita Saikia*

Chemistry Department, New Mexico Highlands University, Las Vegas 87701, USA

Abstract

Trinucleotide repeat expansions underlie several neurodegenerative diseases, including Huntington's disease and other polyQ-related disorders. This work investigates the therapeutic potential of dietary flavonoids, naturally occurring polyphenolic compounds, as candidate small-molecule ligands for recognizing A·A mismatch motifs within CAG trinucleotide repeats and examines how flavonoid chemical structure and electronic descriptors contribute to predicted binding. Fifteen flavonoids representing five subclasses were studied together with naphthyridine-azaquinolone (NA) as a reference ligand using density functional theory, global reactivity descriptors, hydration free-energy analysis, molecular docking, and ADME prediction. NA has been reported to specifically bind slipped-CAG DNA intermediates of expansion mutations. Quercetin and myricetin showed low energy gaps and global hardness comparable to NA, whereas catechin, EGCG, and naringenin demonstrated high global hardness. Docking revealed ligand- and receptor-dependent recognition, with EGCG exhibiting the strongest predicted binding affinity, nobiletin the weakest, and several flavonols displaying favorable interactions across different CAG-repeat structures. Binding strength correlated moderately with chemical potential ($R = 0.62$, $p = 0.011$) and strongly with hydration free energy ($R = -0.82$, $p = 0.0001$). ADME analysis showed that the predicted binding did not necessarily translate into favorable pharmacokinetic behavior; EGCG showed less favorable permeability- and metabolism-related scores, whereas 7,8-dihydroxyflavone, catechin, naringenin, and nobiletin presented balanced final consensus ADME scores. To our knowledge, this work represents one of the few integrated computational studies exploring flavonoids as small-molecule scaffolds for targeting trinucleotide CAG-repeat mismatch motifs.

Corresponding author:

*E-mail: nsaikia@nmhu.edu

ORCID

Nabanita Saikia: 0000-0002-0684-260X

Introduction

Huntington's disease (HD)¹, a rare genetic polyQ disorder, arises from aberrant trinucleotide repeat expansions in exon 1 of the huntingtin gene²⁻⁴. These expansions produce mutant huntingtin protein, which misfolds and aggregates, contributing to neuronal dysfunction⁵. The dynamic instability of trinucleotide repeats is thought to occur via the formation of intrastrand hairpins during replication, repair, recombination, and transcription^{6,7}. Expanded CXXG repeats, where X = A, T, G, or C, form stable RNA and DNA hairpins through G-C base pairing and palindromic CXXG/CXXG motifs. These motifs introduce X·X mismatches, including A·A, C·C, and G·G noncanonical pairs⁸.

Currently available treatments primarily manage symptoms but do not halt disease progression. Hence, understanding the structural effects of repeat expansion is critical for developing targeted interventions that can halt or reverse disease progression^{9,10}. Small molecules offer potential advantages including chemical tunability, ease of administration, and, for suitable compounds, blood-brain-barrier (BBB) permeability^{4,11}. Small molecules can recognize trinucleotide repeat structures, including bulges, internal loops, and hairpins, through hydrogen bonding, π - π stacking, intercalation, electrostatic, van der Waals, and hydrophobic interactions^{1,12,13}. Synthetic compounds, naphthyridine-azaquinolone (NA) and naphthyridine-benzoazaquinolone were shown to exhibit sequence-specific binding to CAG-repeat motifs¹⁴. NA was the first molecule shown to selectively bind to extended CAG hairpins (>30 repeats), forming hydrogen bonds with guanine and adenine while flipping out cytosines from the complementary strand¹⁵. A dataset of 104 flavonoids was examined using ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity), molecular docking, and molecular dynamics simulations on the best conformation from each flavonoid subcategory with the colchicine binding site of tubulin protein¹⁶. The in silico study suggested that flavonoids may act as promising anti-tubulin agents in cancer treatment. A set of flavonoids was docked against inflammation pathway enzymes¹⁷. In the same study, amentoflavone was shown to cause interactions with phosphodiesterase enzymes, while bilobetin and silibinin demonstrated an increase in binding energy with collagenase enzymes. The thermochemical parameters indicated that silibinin was more stable than the other studied flavonoids, whereas ADMET profile predicted that silibinin could be used orally. Isika and Sadik¹⁸ designed, synthesized, and characterized partial flavonoid acetamide derivatives of apigenin, quercetin and luteolin and in vitro bioavailability of the flavonoid acetamide derivatives was determined using the dialysis tubing procedure, while the antioxidant capacity was determined using the 2,2-diphenyl-1-picrylhydrazyl free radical scavenging assay. The structure-activity relationship studies implied that the methylene groups in the flavonoid acetamide derivatives were essential in improving bioavailability, while the hydroxyl group played a key role in the enhancement of antioxidant capacity of flavonoids.

Given the chronic nature of HD treatment, naturally occurring compounds such as dietary flavonoids alongside synthetic ligands may provide complementary chemical scaffolds for developing ligands that selectively recognize repeat-associated nucleic acid structures¹⁹⁻²¹. Dietary flavonoids are naturally occurring, bioactive polyphenols found in various foods and beverages and possess antioxidant, anti-inflammatory, and neuroprotective properties^{22,23}. Flavonoids are a

ubiquitous class of secondary metabolites characterized by a common 15-carbon backbone consisting of two phenyl rings (A and B) linked through a heterocyclic oxygen-containing pyran ring²⁴ (see Figure 1). Based on the chemical structure, flavonoids are categorized into the following subclasses - flavones, flavanols, flavonols, isoflavones, flavanones, and anthocyanins²⁵⁻²⁷.

Flavonoid core scaffold		Flavonoid subclasses included in this study	Flavonoid molecules	Representative structure of each subclass
 ◦ C6–C3–C6 framework ◦ Two aromatic rings, A and B, ◦ Heterocyclic C ring ◦ Subclass differences arise mainly from variations in C-ring unsaturation, oxidation, and hydroxylation.		Flavanols ◦ No C2=C3 double bond ◦ No C4 carbonyl ◦ 3-OH present	EGCG, Catechin	Flavanols
		Flavanones ◦ No C2=C3 double bond ◦ C4 carbonyl present	Hesperetin, Naringenin	Flavanones
		Flavonols ◦ C2=C3 double bond ◦ C4 carbonyl present ◦ 3-OH present	Myricetin, Quercetin, Kaempferol, Fisetin, Morin	Flavonols
		Flavones ◦ C2=C3 double bond ◦ C4 carbonyl present ◦ No 3-OH in core scaffold	Luteolin, Baicalein, Nobiletin, 7,8-DHF, Chrysin	Flavones
		Isoflavones ◦ B ring is attached at C3 instead of C2 ◦ C4 carbonyl present	Genistein	Isoflavones

Figure 1. Overview of the flavonoid core scaffold, flavonoid subclasses, and compounds included in this study. The flavonoid scaffold is shown with the A, B, and C rings labeled. Selected flavonoids are grouped according to subclass, including flavanols, flavanones, flavonols, flavones, and isoflavones.

Flavonoids have attracted research interest in the prevention of various diseases such as cancer, cardiovascular, and neurodegenerative disorders, due to their biochemical, anti-inflammatory²⁸⁻³⁰, antibacterial^{31,32}, anticarcinogenic^{33,34}, and antioxidant (free-radical scavenging) activities^{25,35,36}, which are influenced by the number and position of hydroxyl and other functional groups that govern reactivity, bioavailability, and biological function^{37,38}. Chelation with first-row transition metals in aqueous solution plays an important role in the redox activity of flavonoids, providing protection against oxidative stress through prevention or scavenging of toxic redox-active metal ions^{24,39-41}. Aparicio performed DFT calculations on seventeen flavone derivatives, flavonoids, at the B3LYP/6-311++G** level to investigate the molecular factors that determine the structural and energetic properties in the gas phase²⁶. The study aimed to understand how the presence of hydroxyl groups influenced charge delocalization, and how it was related to geometric and molecular parameters. The torsional profiles around the C2-C1' bond were studied as a function of different hydroxyl group substituents and was related to the charge delocalization. Martins et al.⁴² investigated the relative gas phase acidities of eight flavonoids by means of electrospray-ion trap mass spectrometry. The experimental acidity followed the order, myricetin > luteolin > quercetin > (±)-taxifolin > kaempferol > apigenin > (±)-catechin > (±)-naringenin and showed good agreement with theoretical calculations at the B3LYP/6-311+G(2d,2p)/HF/6-31G(d) level.

For all the flavones and the flavanol catechin, the 4'-hydroxyl group was the most favored deprotonation site whereas for the flavanones, taxifolin and naringenin, the most acidic site was the 7-hydroxyl group. On the other hand, the 5-hydroxyl, in flavones and naringenin, and the 3-hydroxyl, in taxifolin and catechin, were the less acidic positions. The acidity pattern depended on the ortho-catechol group, the 2,3 double bond and the 4-keto group. Myricetin had the highest antioxidant ability, naringenin the lowest, and the others in between. Altunayar-Unsalan and Unsalan⁴³ studied the inhibitory role of plant flavonoid blumeatin toward XO (Xanthine oxidase) and determined its antioxidant mechanism using Fourier transform infrared spectroscopy, DFT, ADME/Tox (absorption, distribution, metabolism, excretion, and toxicity) analysis, and molecular docking. Blumeatin showed better theoretical radical scavenging activity compared to other flavonoids. ADME/Tox analysis determined that blumeatin had a high absorption profile in the human intestine and was neither carcinogenic nor mutagenic⁴³.

The neuroprotective properties of flavonoids in controlling synaptic plasticity and its associated role in brain disorders and treatment of neurodegenerative diseases, including Alzheimer's, Parkinson's, HD, and amyotrophic lateral sclerosis, have been reviewed⁴⁴⁻⁴⁸. Myricetin was reported to reduce mutant huntingtin expression and prevent sequestration of muscle blind like-1 splicing factor through interaction with expanded CAG-repeat RNA⁴⁹. Structural analysis of myricetin-5'CAG/3'GAC RNA complex indicated that myricetin binds near the A·A mismatch motif, while cellular experiments showed reduced mutant huntingtin aggregation. A prior study reported sequence-dependent binding of flavonoids, quercetin and flavopiridol to DNA duplexes, supporting their potential as RNA-targeting molecules⁵⁰. Flavonoid, naringin attenuated the 3-Nitropropionic acid (3-NP)-induced neurodegeneration through the modulation of intrinsic apoptotic cascade in Wistar rats⁵¹. The study showed that in rats exposed to 3-NP, naringin treatment at 80 mg/kg body weight reduced lipid peroxidation and protein carbonyl levels, restored enzymatic and non-enzymatic antioxidant defenses, and normalized striatal ATPase activities, indicating protection against 3-NP-induced neuronal apoptosis via decreased cytochrome c release from mitochondria and caspase 3 activation as revealed by Western blot.

Despite growing evidence supporting the neuroprotective potential of flavonoids, the molecular-level basis underlying recognition of CAG-repeat trinucleotide motifs in DNA and RNA remains poorly understood, and studies in this research area are limited. In particular, the relationship between flavonoid electronic structure, global reactivity descriptors, pharmacokinetic properties, and interactions with mismatch motifs within trinucleotide-repeat-containing DNA/RNA structures has not been systematically investigated, in detail, at the molecular level. Because ligand binding to nucleic acid structures depends on hydrogen bonding, aromatic π -stacking, electrostatic interactions, molecular polarity, and sequence context, quantum chemical descriptors are valuable for identifying the physicochemical factors that may govern flavonoid-nucleic-acid recognition.

To address this knowledge gap, we employed an integrated computational framework combining DFT-based electronic structure analysis, global reactivity descriptors, pharmacokinetic

profiling, and molecular docking. Fifteen dietary flavonoids associated with antioxidant or neuroprotective activity in HD and related neurodegenerative disorders were selected from prior reports^{22,53-55}, as shown in Figure 1. The synthetic ligand NA was included as a reference compound for comparison. Molecular docking of the selected flavonoids and NA with NMR and X-ray crystal structures of RNA and DNA containing varying CAG-repeat expansions with A·A mismatches was performed to investigate the binding modes within the noncanonical mismatch motifs. Docking results were interpreted alongside DFT-derived electronic descriptors and predicted pharmacokinetic properties to examine whether these properties correlate with any known or predicted CAG-repeat recognition.

Methods

Compound dataset and selection criteria

The flavonoid dataset included representatives from five major flavonoid subclasses: **flavanols** (catechin and epigallocatechin gallate, EGCG), **flavones** (chrysin, luteolin, baicalein, nobiletin, and 7,8-dihydroxyflavone, 7,8-DHF), **flavonols** (quercetin, myricetin, kaempferol, fisetin, and morin), **flavanones** (hesperetin and naringenin), and the **isoflavone** genistein (see Figure S1a of Supplementary Information). The synthetic CAG-repeat-targeting ligand NA was included as a reference compound for comparison (see Figure S1b of Supplementary Information). Structural diversity was used as a selection criterion to ensure a broad yet manageable coverage of flavonoid subclasses, hydroxylation patterns, and functional group distributions within the molecular scaffolds.

Ligand preparation and quantum chemical calculations

The structures of the fifteen flavonoids and NA were obtained from PubChem⁵⁶. Initial molecular geometries were prepared using GaussView 6.1.1 and geometries were optimized using Gaussian 16 software⁵⁷. GaussView, Avogadro, and ChemCraft software were used for electronic structure analysis and visualization. Geometry optimizations were performed in the gas phase and in implicit solvent using the M06-2X and CAM-B3LYP functionals with the 6-31G(d,p) and 6-31+G(d,p) basis sets, respectively. The CAM-B3LYP Coulomb-attenuated functional was selected because it has been reported to provide consistent performance for electronic-structure calculations involving excitation-energy^{52,58}. Vibrational frequency calculations were performed on all optimized geometries to confirm the absence of imaginary frequencies and verify that the optimized structures correspond to local minima on the potential energy surface. The basis sets were chosen based on a previous study that reported HOMO–LUMO energy gap values to converge above the 6-31G(d) basis set, with an absolute deviation of <1% (~0.03 eV) relative to the 6-311++G(d,p) basis set⁵⁹. Increasing the basis set size from 6-31G to 6-311G or from 6-31G(d) to 6-311G(d) had limited improvement in the computed energy gaps.

To benchmark the DFT-derived electronic energies, single point calculations were performed using the domain-based local pair natural orbital coupled-cluster method with single, double, and perturbative triple excitations, DLPNO-CCSD(T)^{60,61,62}. The high-level calculations were carried out at the DLPNO-CCSD(T)/def2-TZVP and DLPNO-CCSD(T)/cc-pVTZ using the ORCA package with NormalPNO parameters and TightSCF convergence criteria^{60,63,64}. The higher level

of theory was selected based on a prior work on noncanonical adenine base pair stability and related nucleic acid systems⁶⁵. Solvent phase calculations were performed using the Solvation Model based on Density (SMD)⁶⁶. Three solvents were considered: water, ethanol, and dimethyl sulfoxide (DMSO), with dielectric constants of 80.1, 24.5, and 46.7, respectively. The three solvents were categorically selected to compare the effects of polar protic, moderately polar protic, and polar aprotic environments on the electronic structure and solvation behavior of flavonoids and the reference ligand NA.

Frontier molecular orbitals and global reactivity descriptors

Frontier molecular orbital analysis was performed using optimized geometries. The highest occupied molecular orbital energy (E_{HOMO}), lowest unoccupied molecular orbital energy (E_{LUMO}), and HOMO–LUMO energy gap (ΔE) were calculated to evaluate electronic stability and chemical reactivity. The energy gap was calculated as:

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

where, E_{HOMO} characterizes the electron donating while E_{LUMO} characterizes the electron withdrawing ability of a molecule. A small energy gap suggests ease in electron transfer from HOMO to LUMO states, which makes a molecule chemically reactive and less stable whereas, a large energy gap indicates that more energy is needed for electronic transitions, meaning the molecule is relatively stable or less reactive. Further, localization of electron density in HOMO highlights nucleophilic regions, whereas localization in the LUMO highlights electrophilic regions⁶⁷. The ionization potential (IP) associated with the HOMO indicates the energy required to remove an electron while the electron affinity (EA), related to the LUMO represents the energy released when capturing an electron⁶⁷.

$$\begin{aligned} IP &= -E_{\text{HOMO}} \\ EA &= -E_{\text{LUMO}} \end{aligned} \quad (2)$$

Global reactivity descriptors, including chemical potential (μ), electronegativity (χ), global hardness (η), global softness (S), electrophilicity index (ω), and maximum charge transfer index (ΔN_{max}) were calculated using Koopmans' theorem, from the IP and EA (frontier molecular orbital energies) values. These descriptors were used to compare the relative stability/reactivity, charge transfer, and electrophilic (nucleophilic) character of the flavonoids and NA. The quantum molecular descriptors derived from the E_{HOMO} and E_{LUMO} and IP and EA energies are provided in Table 1. Details of the quantum molecular descriptors can be found elsewhere⁶⁸⁻⁷². The maximum amount of electron charge transfer from the nucleophile to the electrophile, ΔN_{max} , is given in Table 1. Larger ΔN_{max} values indicate a greater electron accepting capacity and stronger potential for charge transfer interactions, whereas smaller values suggest a lower tendency to accept electron density.

Table 1. Quantum molecular descriptors and the corresponding equations used for the conceptual DFT analysis in this study.

Descriptor		Equation
HOMO-LUMO energy gap	ΔE	$E_{\text{LUMO}} - E_{\text{HOMO}}$
Chemical potential	μ	$\mu = -\frac{(IP + EA)}{2}$
Electronegativity	χ	$\chi = \frac{IP + EA}{2}$
Global hardness	η	$\eta = \frac{IP - EA}{2}$
Global softness	S	$S = \frac{1}{\eta}$
Electrophilicity index	ω	$\omega = \frac{\mu^2}{2\eta}$
Max. charge transfer index	ΔN_{max}	$\Delta N_{\text{max}} = -\frac{\mu}{\eta}$

Hydration free energy

Hydration free energy (ΔG_{hyd}) was calculated as the free energy change associated with transferring a molecule from gas phase to aqueous solution and is given as:

$$\Delta G_{\text{hyd}} = \Delta G_{\text{water}} - \Delta G_{\text{gas}} \quad (3)$$

Negative ΔG_{hyd} ($\Delta G_{\text{hyd}} < 0$) values were interpreted as thermodynamically favorable hydration. In this study, dipole moment and ΔG_{hyd} are used as important molecular descriptors in defining regions of chemical space for flavonoid and NA binding.

In silico ADME and toxicity prediction

The pharmacokinetic and toxicological profiles of the selected flavonoids and NA were evaluated using computational ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) prediction tools^{73,74}, including SwissADME⁷⁵, ADMETlab3⁷⁶, and admetSAR3.0⁷⁷ web servers to discern whether our studied compounds have the potential for any adverse effect in the human body upon administration. Canonical SMILES (simplified molecular-input line-entry system) strings were used as input structures for all web-based prediction platforms. The evaluated pharmacokinetic parameters included gastrointestinal absorption, Caco-2 permeability, BBB permeation, plasma protein binding (PPB), cytochrome P450 (CYP) metabolism, and general drug-likeness properties^{78,79}. Drug-likeness was assessed using Lipinski's Rule of Five, including molecular weight, lipophilicity (logP), number of hydrogen-bond donors, and number of

hydrogen-bond acceptors. Predicted ADME and toxicity parameters were compared to identify compounds with favorable pharmacokinetic profiles and reduced predicted toxicity.

Molecular Docking

Molecular docking was performed to investigate the predicted binding modes and binding affinities of the selected flavonoids and the reference ligand NA toward representative CAG-repeat-containing RNA and DNA structures with varying noncanonical A·A mismatch motifs. The three-dimensional nucleic acid structures were obtained from the Protein Data Bank (PDB). Table 2 lists the PDB IDs, classification, and experimental method of the available CAG-repeat-containing RNA and DNA structures.

Table 2. The crystal structures of CAG-repeat expansions retrieved from the PDB that are used for molecular docking. Here, ‘*n*’ represents the number of A·A mismatch repeats within the CAG structure. CAG-repeat RNA and DNA motifs were selected to compare ligand recognition of repeat-associated noncanonical base pairs.

<i>n</i>	PDB ID	Method	Classification
1	5X11, 9DLL, 9DKG, 9DLM, 9DLN, 2MS5, 7D12, 8ZNQ	Solution NMR	RNA
	3NJ6 (3NJ7)	X-ray diffraction	RNA
2	9KAD	Solution NMR	RNA
3	5VH7	Solution NMR	RNA
	4YN6, 4J50	X-ray diffraction	RNA
5	7VFT	X-ray diffraction	RNA
1	1X26	Solution NMR	DNA
3	8WU5	X-ray diffraction	DNA

For each repeat length, *n*, one representative structure was selected from the PDB. Structures were chosen based on experimental method and structural resolution, presence of the target CAG-repeat motif, and suitability for docking preparation. This approach was used to enable consistent comparison of predicted binding modes across the reference ligand NA and flavonoid dataset. Prior to docking, receptors were prepared by removing any crystallographic water molecules, ions, and non-relevant ligands where appropriate. Hydrogen atoms were added, and receptor files were converted into the required docking format using Chimera⁸⁰ software. Ligand structures were obtained from the gas phase optimized geometries from quantum chemical calculations. Ligand rotatable bonds were defined prior to docking. Both ligands and receptors were converted to “pdbqt” file format in Chimera to prepare the structures for docking.

Docking calculations were carried out using AutoDock Vina⁸¹. The docking grid was centered around the noncanonical A·A mismatch region of each CAG-repeat nucleic-acid structure. Depending on the crystal structure of the receptors, grid box dimensions were varied and an exhaustiveness value of 32 was used, and 20 poses were considered. Details of the docking scoring function can be found elsewhere⁸¹. For each ligand/receptor pair, the top-ranked docking poses were evaluated based on predicted binding affinity, orientation within the CAG-repeat motif,

proximity to A·A mismatch sites, hydrogen-bonding, π - π stacking, electrostatic contacts, and hydrophobic interactions. The docking score was treated as the predicted binding affinity and expressed in kcal/mol. More negative values were interpreted as stronger predicted ligand–nucleic acid binding. Docked poses were visualized and analyzed using PyMOL and Chimera software. Docking results were interpreted alongside quantum chemical descriptors and ADMET predictions to elucidate whether electronic structure, polarity, solvation, and pharmacokinetic properties correlated with predicted CAG-repeat recognition. Where a co-crystallized or experimentally characterized ligand was available, the docking protocol was validated by redocking the reference ligand and comparing the predicted pose with the experimental binding orientation.

The plots were generated using Python with pandas, NumPy, and Matplotlib. AutoDock Vina predicted binding affinities were organized into CSV files before plotting. Ligands were grouped according to flavonoid subclass, with NA included as the reference ligand. Docking heatmaps were generated using AutoDock Vina predicted binding affinities reported in kcal/mol, with more negative values indicating stronger predicted binding. Representative docking poses were visualized using molecular-graphics software, and selected poses were prepared to show ligand orientation and nearby interacting nucleobases. OriginPro and Pymol were also used for generating the figures.

Results and Discussion

Benchmark of DFT levels of theory

The optimized geometries of the flavonoids and NA are shown in Figure S1 of Supplementary Information. To evaluate the reliability of selected DFT methods in describing the electronic properties and quantum molecular descriptors of flavonoids dataset and reference ligand NA, a benchmark study was performed using the HOMO–LUMO energy gap (ΔE) as a key electronic descriptor. The performance of M06-2X and CAM-B3LYP functionals, in combination with different basis sets, was compared against DLPNO-CCSD(T), which was used as a higher-level correlated reference method⁸². The DLPNO-CCSD(T) method is a local correlation method based on pair natural orbitals (PNO) that assumes that the electron correlation is a local phenomenon and decays rapidly with an increase in the inter-electronic distance. Compared to the traditional CCSD, DLPNO-CCSD(T) produces reasonable CCSD(T) results with reduced computational cost and can be extended to large molecular systems.

As shown in Figure 2a, both M06-2X and CAM-B3LYP underestimated the energy gap relative to DLPNO-CCSD(T) in implicit waters. This difference is expected because the DLPNO-CCSD(T) calculated values were obtained from correlated total energy differences⁸², whereas the DFT-derived values represent Kohn–Sham orbital energy differences. However, both functionals successfully reproduced the relative trends in energy gaps across the different flavonoid subclasses. Less conjugated structures such as flavanols and flavanones showed higher energy gaps whereas flavonols, flavones, and isoflavones along with the synthetic compound NA exhibited lower energy gap values (see Figure 2a). The observed trend can be correlated to the degree of π -conjugation in compounds containing the C2=C3 double bond, C4 carbonyl group, and molecular planarity: conjugated systems, such as flavonols, flavones, and isoflavones

stabilized the LUMO and reduced the energy gap. A comparatively lower energy gap in NA supports its electronically soft behavior relative to flavonoids.

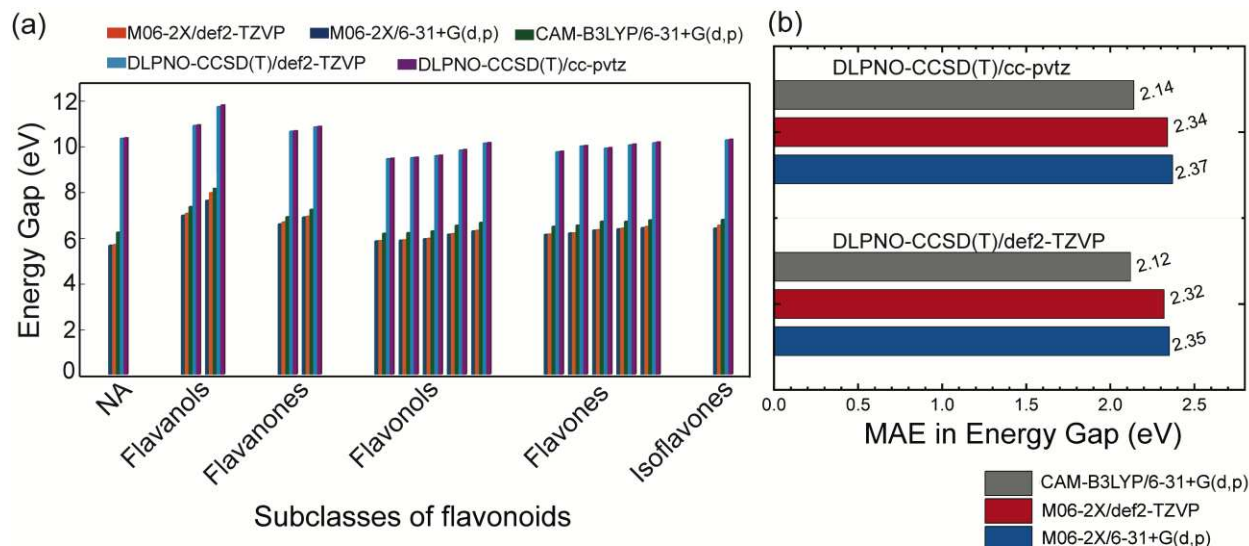


Figure 2. (a) The performance of the functionals M06-2X and CAM-B3LYP, in combination with different basis sets, against DLPNO-CCSD(T) level in presence of implicit water. (b) Mean absolute error (MAE) in energy gap of M06-2X and CAM-B3LYP, relative to DLPNO-CCSD(T).

The mean absolute error (MAE), a statistical metric used to assess the average magnitude of error between calculated and reference values, is employed to quantify the error in the calculated energy gap, ΔE . MAE is calculated using the equation:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

where, y_i and $\hat{y}_i$ are the ΔE values calculated for different DFT functionals, and ‘ n ’ is the number of observables. To quantify the accuracy of the two functionals, the MAE in energy gap was calculated and the results were benchmarked against DLPNO-CCSD(T) level, as shown in Figure 2b. When benchmarked against DLPNO-CCSD(T)/cc-pVTZ, CAM-B3LYP/6-31+G(d,p) showed a MAE of 2.14 eV, while M06-2X/def2-TZVP and M06-2X/6-31+G(d,p) yielded MAE values of 2.34 and 2.37 eV, respectively. Similarly, when compared to DLPNO-CCSD(T)/def2-TZVP, CAM-B3LYP/6-31+G(d,p) exhibited an MAE of 2.12 eV, while M06-2X/def2-TZVP and M06-2X/6-31+G(d,p) had values of 2.32 and 2.35 eV, respectively. CAM-B3LYP showed slightly lower MAE values and better agreement with the DLPNO-CCSD(T) reference method in predicting the energy gap values.

Dipole Moment

The dipole moment was computed to understand solvent-dependent changes in polarity of the flavonoids and the reference ligand NA, as well as to assess how electronic charge distribution influences interaction potential. As shown in Figure 3, dipole moments were calculated in both the gas and implicit solvent phase at the M06-2X/6-31+G(d,p) and CAM-B3LYP/6-31+G(d,p) levels.

For each compound, the connecting line between the gas and solvent phase values represents the solvent-induced change in dipole moment.

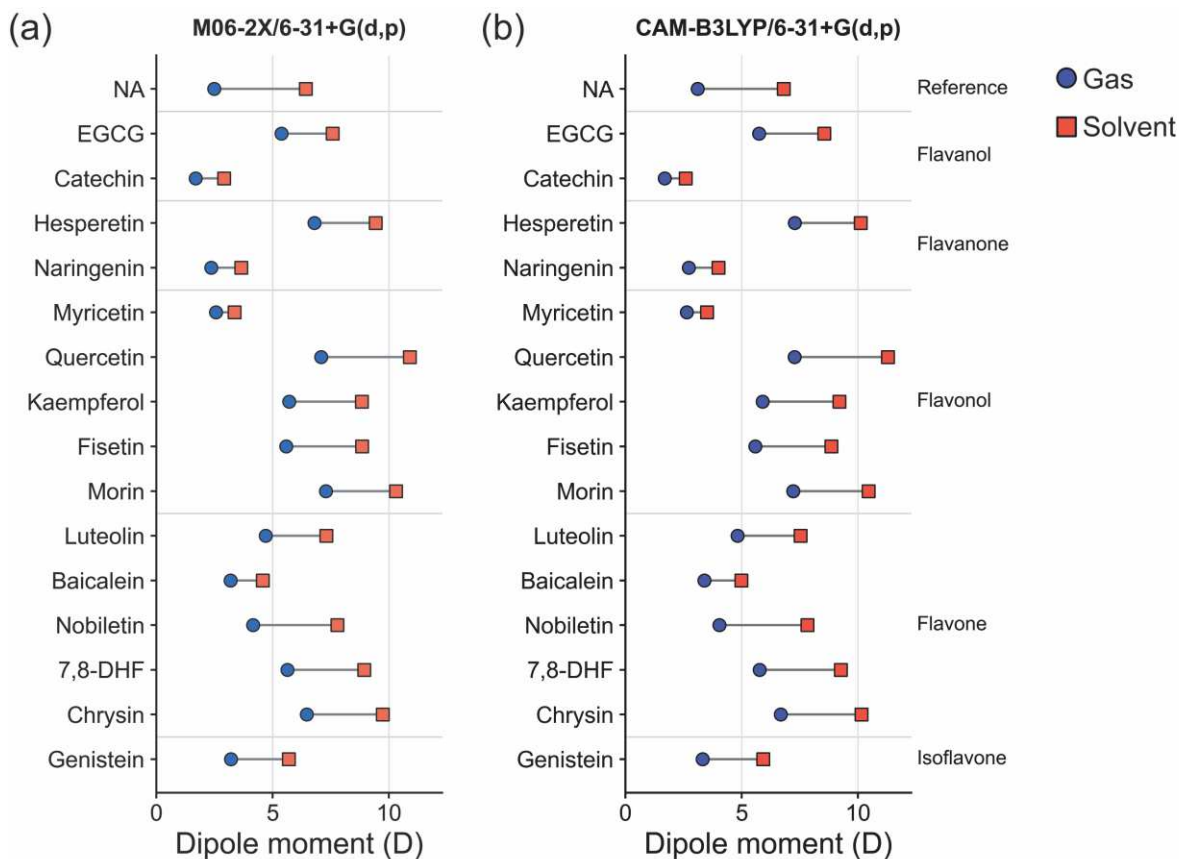


Figure 3. Dipole moments of NA and the flavonoid dataset calculated in the gas phase and implicit solvent phase at the (a) M06-2X/6-31+G(d,p) and (b) CAM-B3LYP/6-31+G(d,p) levels of theory. Compounds are grouped according to flavonoid subclass, with NA included as a reference ligand. The connecting line between the gas and implicit solvent phase markers represents the magnitude of the solvent induced increase in molecular polarity. For all compounds, the dipole moment increased in implicit water relative to the gas phase, indicating enhanced molecular polarization in solvent phase. The trends are comparable between the two levels of theory, with some compounds showing stronger solvent-dependent changes than others.

The higher dipole moments observed in the solvent phase relative to gas phase indicate enhanced polarization in the aqueous environment, due to solvent induced stabilization of charge separated electronic distributions, particularly in compounds containing polar hydroxyl and carbonyl functional groups. Flavonoids such as quercetin, morin, hesperetin, chrysin, kaempferol, fisetin, 7,8-DHF, and EGCG presented comparatively high dipole moments, suggesting greater polarity in aqueous implicit solvent. In contrast, catechin, naringenin, and myricetin had lower dipole moments across both phases and levels of theory. The ligand NA exhibited a considerable increase in dipole moment upon solvation, highlighting the importance of aqueous environment in modulating its polarity. The observed variations in dipole moments are linked to differences in

molecular geometry, electronic distribution, position and orientation of hydroxyl groups, and extent of π -conjugation.

Flavonoids exhibit electron delocalization across the A, B, and C rings, and the electronic properties are influenced by the number and position of hydroxyl groups^{26,83}. Molecules with asymmetric structures and uneven distribution of electronegative atoms exhibit larger dipole moments because of greater separation of partial charges. A high dipole moment of quercetin may arise from its geometry, particularly rotation of the B ring relative to the A-C ring⁸⁴, which disrupts molecular symmetry and leads to an uneven distribution of electron density⁸⁵. In addition, the hydroxyl group arrangement in quercetin may enhance localized electronegativity differences, resulting in a larger net dipole moment. A similar behavior is observed in morin, where the asymmetric hydroxyl group substitution pattern contributes to charge separation.

In contrast, the lower dipole moment of myricetin may be associated with a balanced distribution of hydroxyl and oxo functional groups and a greater degree of charge compensation across the molecule. This interpretation is consistent with previous computational studies showing that flavonoids containing a 3-OH group can adopt relatively planar conformations through intramolecular hydrogen bonding, which enhances π -delocalization across the ring system^{26,43}. Increased molecular planarity can promote an even charge distribution, leading to partial cancellation of local dipole vectors and a lower dipole moment. A higher dipole moment of chrysin may reflect the combined effects of its carbonyl group, hydroxyl substitution, and molecular asymmetry, despite the absence of hydroxyl groups on the B ring. By contrast, catechin exhibited a relatively low dipole moment, which may be attributed to its non-planar, saturated C-ring structure that disrupts extended π -conjugation.

To further quantify the effect of implicit waters, the solvent-induced change in dipole moment was calculated as ($\Delta\mu = \mu_{\text{water}} - \mu_{\text{gas}}$) and summarized in Table S1 of Supplementary Information. All compounds exhibited positive $\Delta\mu$ values at both levels of theory, confirming that implicit water increases molecular polarization. The magnitude of $\Delta\mu$ varied substantially among the dataset, indicating that solvent sensitivity depended on molecular geometry, substitution pattern, and electronic distribution rather than the number of polar functional groups. NA showed a pronounced solvent-induced increase in dipole moment, consistent with its molecular geometry that facilitates strong polarization in aqueous solvent. Among the flavonoids, quercetin, nobiletin, 7,8-DHF, chrysin, kaempferol, fisetin, and morin showed relatively large $\Delta\mu$ values, suggesting greater sensitivity to aqueous solvation. In contrast, catechin and myricetin had smaller $\Delta\mu$ values, indicating comparatively weaker solvent induced polarization. The π -delocalization across the flavonoid scaffold plays an important role in determining electronic distribution and molecular polarity, whereas disruption of conjugation can reduce resonance-driven charge separation.

Dipole moment in different solvents

To further investigate solvent-dependent polarization, dipole moments were evaluated in water, ethanol, and DMSO. The solvents were selected to compare environments with different dielectric constants and protic/aprotic character. Water, with the highest dielectric constant, ϵ of 80.1, represents a highly polar aqueous environment relevant to physiological conditions. Ethanol, $\epsilon =$

24.5 and DMSO, $\epsilon = 46.7$ were included as commonly used polar solvents in spectroscopic and biochemical studies, providing additional comparison of protic and aprotic solvent effects. Across all flavonoids and NA, the calculated dipole moments generally followed the trend: water > ethanol > DMSO as shown in Figure S2 of the Supplementary Information at the M06-2X/6-31+G(d,p) level. Most ligands exhibited higher dipole moments in water than in ethanol or DMSO, consistent with stronger solvent induced polarization in the aqueous implicit-solvent environment. Because the SMD model treats the solvent implicitly, these observed differences should be interpreted primarily in terms of bulk solvation effects rather than specific, explicitly modeled solute–solvent hydrogen bonds. Nevertheless, the hydroxyl and carbonyl groups present in flavonoids and reference ligand NA influence the intrinsic polarity and enhance sensitivity to polar solvent environments. Ethanol, as a polar protic solvent, supported stabilization of polar electronic distributions, although to a lesser extent compared to water. DMSO, despite its relatively high dielectric constant, showed lower calculated dipole moments, suggesting that dielectric constant alone does not fully describe the solvent-dependent polarization behavior of these compounds. The solvent effects on dipole moment depend on bulk solvent properties and molecular features such as orientation and distribution of functional groups, electronic structure, and scaffold geometry. Subsequent comparisons therefore focus on the aqueous implicit solvent results.

Energy gap and global hardness: Influence on reactivity and molecular stability

The HOMO–LUMO energy gap (ΔE) and global hardness (η) were analyzed to evaluate the electronic stability and chemical reactivity of the flavonoids and reference ligand NA in implicit solvent, as shown in Figure 4. The energy gap and η are closely associated with molecular reactivity. A smaller energy gap and lower η value indicate higher chemical reactivity and greater ability to participate in charge transfer interactions, which may be relevant to noncanonical mismatch recognition in CAG-repeat structures. A consistent trend is observed across both functionals (M06-2X and CAM-B3LYP), with M06-2X generally predicting slightly lower energy gaps.

Quercetin and myricetin exhibited the lowest energy gaps and η values, comparable to the reference molecule NA, suggesting that they are among the electronically softer compounds in the dataset and may readily undergo charge redistribution during noncovalent interactions. For quercetin and myricetin, this may be associated with extended π -conjugation across the A-C-B ring system, facilitated by the presence of the C2=C3 double bond. The extended conjugation stabilizes the frontier molecular orbitals and reduces the energy gap, thereby enhancing reactivity. The other flavonoids exhibited moderate and broadly comparable energy gap and η values, indicating similar electronic behavior across much of the flavonoid dataset. However, a significant deviation was observed in catechin, EGCG, and naringenin, which exhibited higher energy gaps and η values, indicating lower reactivity (see Figure 5).

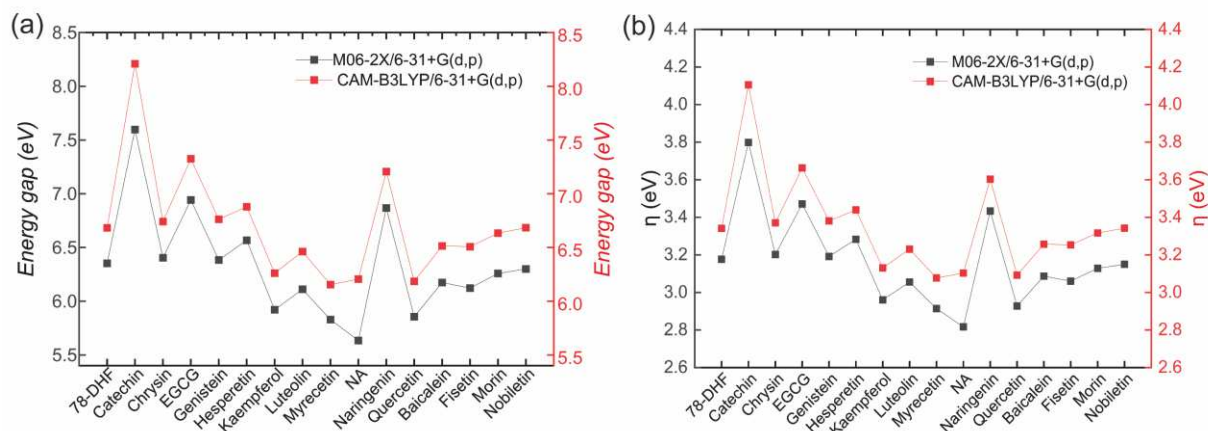


Figure 4. (a) Energy gap (ΔE) and (b) global hardness (η) of flavonoids and reference ligand NA in implicit solvent phase.

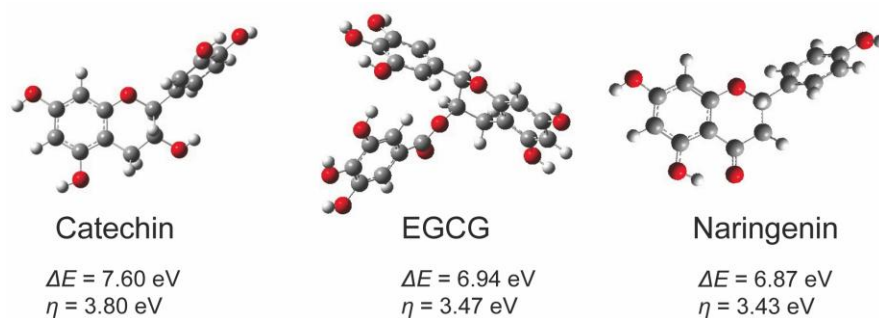


Figure 5. The energy gap and global hardness (η) values of catechin, EGCG, and naringenin.

The higher energy gap and η observed in catechin (Figure 5, left panel) is attributed to its saturated C2–C3 bond, which disrupts π -conjugation across the flavonoid backbone and limits electron delocalization. Similarly, naringenin, which lacks the C2=C3 bond, exhibits a comparable increase in energy gap and η value due to reduced conjugation (Figure 5, right panel). EGCG, with a non-planar, conformationally flexible structure, behaves similarly to catechin, which further disrupts π -conjugation across the molecule, as shown in Figure 5, middle panel. Despite having multiple hydroxyl groups, lack of planarity and disruption of π -conjugation reduce electron delocalization, resulting in relatively higher energy gap and η values. The ΔE and η analysis indicates that electronic reactivity is governed not only by the number of hydroxyl groups, but also by molecular planarity, conjugation, scaffold geometry, and the ability of the molecule to delocalize electron density.

Chemical potential and global softness

The chemical potential, μ , is a useful conceptual DFT descriptor that reflects the tendency of a molecule to lose electron density. In general, a more negative μ value indicates a lower escaping tendency of electrons and therefore greater electronic stability, whereas a less negative μ value suggests a higher tendency to donate electrons and relatively greater reactivity. Because chemical potential is related to electronegativity ($\mu = -\chi$), less negative μ values indicate a greater relative

tendency toward electron-density donation, whereas more negative values indicate stronger retention of electron density. The μ values are negative, indicating a thermodynamically favorable and stable electronic configuration that resists decomposition or spontaneous chemical change^{86,87}. As shown in Table S2 of Supplementary Information, at the M06-2X/6-31+G(d,p) level, catechin has the least negative μ value of -3.544 eV, followed by NA (-3.827 eV) suggesting a relatively greater tendency toward electron-density donation. A more negative μ value for nobiletin (-4.485 eV) and chrysin (-4.473 eV) indicates the ligands to be electronically more stable or poor electron donors. The variation in μ among the compounds reflects differences in the electron-donating tendency, which may influence electronic interactions with CAG-repeat-containing nucleic-acid targets.

Chemical softness (S) was evaluated as an additional conceptual DFT descriptor to assess the relative electronic reactivity of the flavonoids and the reference ligand NA (Table S2, Supplementary Information). Global softness is inversely related to global hardness (see Table 1); a higher S value indicates greater electronic softness, lower resistance to electron-density redistribution, and higher relative chemical reactivity. In contrast, lower S value indicates greater electronic hardness and comparatively lower reactivity. The decreasing order of S was, NA > myricetin > quercetin > kaempferol > luteolin > fisetin > baicalein > morin > nobiletin > 7,8-DHF > genistein > chrysin > hesperetin > naringenin > EGCG > catechin. NA, myricetin, and quercetin are among the electronically reactive compounds in the dataset, consistent with their relatively low η values and enhanced ability to participate in charge transfer or polarization driven interactions. In contrast, EGCG and catechin showed the lowest global softness values, suggesting greater electronic hardness and lower electronic reactivity. The global softness trend supports the energy gap and global hardness analyses, suggesting that differences in π -conjugation, molecular planarity, and electron delocalization strongly influence the reactivity of flavonoid scaffolds and reference ligand NA.

The maximum charge transfer index ($\Delta N_{\max}$) was evaluated to compare the relative electron-accepting capacity of NA and the flavonoid dataset (see Table S3 of Supplementary Information). At the M06-2X level, most compounds showed comparable $\Delta N_{\max}$ values, with quercetin, nobiletin, baicalein, kaempferol, EGCG, and myricetin exhibiting higher values. NA also fell within the comparable range, suggesting that several flavonoids possess charge-transfer capacities similar to the reference ligand. The CAM-B3LYP results showed a similar pattern for most compounds, although catechin displayed a markedly lower value at both levels. Overall, the $\Delta N_{\max}$ analysis suggests that several flavonoid scaffolds, particularly flavonols and selected flavones, may support charge-transfer upon binding with nucleobases in CAG-repeat structures.

Electronegativity and electrophilicity index

The electronegativity (χ) and electrophilicity index (ω) were calculated at the M06-2X/6-31+G(d,p) and CAM-B3LYP/6-31+G(d,p) levels of theory, as shown in Figure 6. The two global reactivity descriptors provide insight into the electron-accepting ability and electrophilic character of the compounds. The ω describes the stabilization energy of a molecule when it receives electrons and represents electrophilicity, with larger values of ω indicating stronger electrophilic

behavior⁸⁸. Higher χ and ω values generally indicate a greater tendency to accept electron density⁸⁹, which may influence interactions with nucleobases. Since a molecule with good electrophilicity has a low μ value⁹⁰, at the M06-2X level, most flavonoids, except NA, EGCG, and catechin, exhibited comparable χ values in the range of ~ 3.5 - 4.5 eV and ω values ~ 2.5 - 3.2 eV, indicating similar electrophilic behavior.

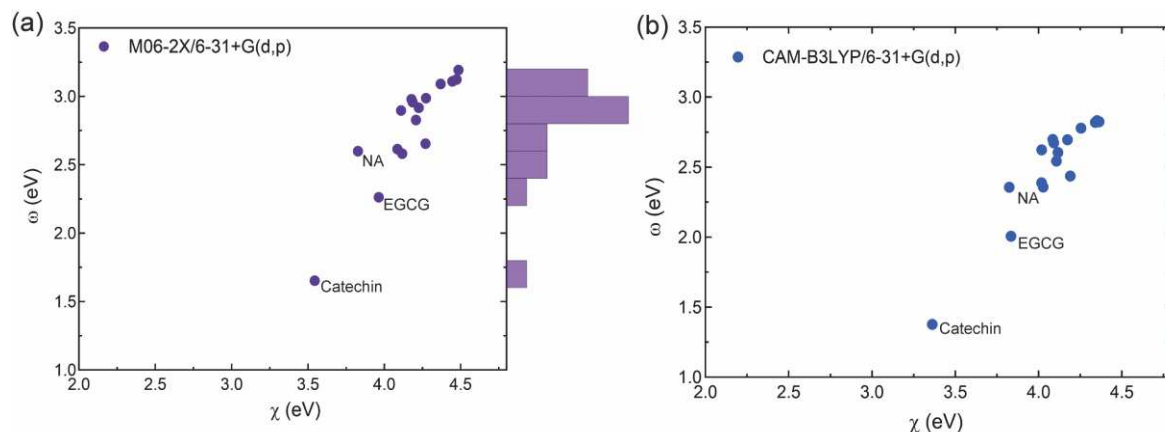


Figure 6. Electronegativity (χ) versus electrophilicity index (ω) of the reference ligand NA and selected flavonoids calculated in aqueous implicit solvent at the (a) M06-2X/6-31+G(d,p) and (b) CAM-B3LYP/6-31+G(d,p) levels of theory. Histograms show the distribution of ω values, and selected compounds are labeled to highlight the deviations in electronic descriptors.

A similar trend was observed at the CAM-B3LYP level, where most flavonoids clustered within a narrower region, with χ values of ~ 3.8 - 4.5 eV and ω values of ~ 2.3 - 2.8 eV. Catechin was an outlier with significantly lower χ and ω values, while EGCG also deviated from the main cluster, although to a lesser extent. In catechin, torsional rotation of the B ring relative to the A-C ring system disrupts π -conjugation, contributing to its lower electrophilicity. Catechin also lacks the C2=C3 double bond and C4 carbonyl group required for extended π -conjugation. Consequently, catechin exhibits lower electron affinity and reduced electrophilic character. Despite the presence of multiple hydroxyl groups, EGCG showed reduced electrophilicity, likely because its nonplanar and conformationally flexible structure limits π -delocalization and electron-accepting character. Flavonoids such as quercetin, myricetin, luteolin, and morin showed moderate to high χ and ω values, reflecting higher electrophilic character due to the presence of C2=C3 double bond, the C4 carbonyl group, and extended π -conjugation, which promote charge delocalization across the flavonoid scaffold. The χ - ω analysis showed that electronic descriptors were strongly influenced by conjugation and substitution patterns within the molecule.

Hydration free energy

Hydration free energy (ΔG_{hyd}) is defined as the change in Gibbs free energy associated with transferring a molecule from the gas to aqueous phase. It is an important thermodynamic descriptor related to aqueous stabilization, solvation, partitioning, and bioavailability of small molecules, making it relevant in drug design^{91,92}. In general, a more negative ΔG_{hyd} indicates favorable hydration and stronger stabilization in water. Figure S3 of Supplementary Information depicts the

ΔG_{hyd} of the flavonoids and the reference ligand NA at the two levels of theory, M06-2X and CAM-B3LYP. Both functionals showed consistent trends, although NA and EGCG displayed notable method-dependent differences relative to the remaining compounds.

Flavonoids were categorized into strong hydration (EGCG, myricetin, catechin, fisetin), moderate hydration (quercetin, luteolin, naringenin, hesperetin, genistein, kaempferol), and weak hydration (7,8-DHF, chrysin, baicalein, nobiletin). NA exhibited the most favorable ΔG_{hyd} followed by EGCG, while the remaining flavonoids were within the moderate-to-weak range. A more favorable hydration behavior of EGCG was attributed to the increased number of hydroxyl groups and presence of a gallate moiety, which increased polarity, intermolecular hydrogen bonding, and favored aqueous stabilization. Flavanols and flavonols exhibited more negative to moderately negative ΔG_{hyd} values, reflecting their higher hydroxyl content and greater hydrogen bonding capacity. This trend suggests increased aqueous stabilization for highly hydroxylated flavonoids, although favorable hydration alone does not necessarily imply stronger target binding because desolvation effects and binding site properties may influence ligand recognition. Although favorable hydration can increase aqueous stabilization, excessive solvation may reduce the availability of a ligand for direct interaction with nucleic-acid binding sites.

Pharmacokinetic properties using in silico ADME and toxicity

To further evaluate the drug-likeness and pharmacokinetic properties of the selected flavonoids and reference ligand NA, computational ADME and toxicity predictions were performed. The analysis included absorption-, distribution-, metabolism-, excretion-, and toxicity-related descriptors, including gastrointestinal absorption (HIA), Caco-2 permeability, BBB permeability, plasma protein binding (PPB), cytochrome P450 metabolism, Lipinski's Rule of Five parameters, and selected toxicity endpoints. The selected properties were used to assess whether the compounds with favorable electronic descriptors and docking behavior (discussed in the next section) also possess acceptable predicted pharmacokinetic profiles. The predicted pharmacokinetic profiles are summarized using a consensus ADME heatmap as shown in Figure 7.

Raw outputs from ADMETlab3 and admetSAR3.0 tools were converted into coded values, where scores of **2**, **1**, and **0** represented favorable, intermediate, and unfavorable predicted behavior. The scoring approach allowed comparison of ADME properties that were originally reported on different numerical or categorical scales. The heatmap was divided into individual ADME parameters and category-level scores to distinguish specific pharmacokinetic features from ADME suitability. The individual descriptor heatmap showed that most compounds satisfied essential physicochemical and drug-likeness criteria, including acceptable molecular weight (MW), lipophilicity, hydrogen-bond donor/acceptor counts, and Lipinski rule. In contrast, greater variation was observed for TPSA (Topological Polar Surface Area), Caco-2 permeability, BBB permeability, PPB, and CYP inhibition risk. The predicted pharmacokinetic limitations of the flavonoid dataset are associated primarily with permeability, distribution, and metabolism rather than drug-likeness. NA showed moderate final consensus ADME score, with favorable molecular weight, logP, Lipinski rule, solubility, and PPB-related scores.

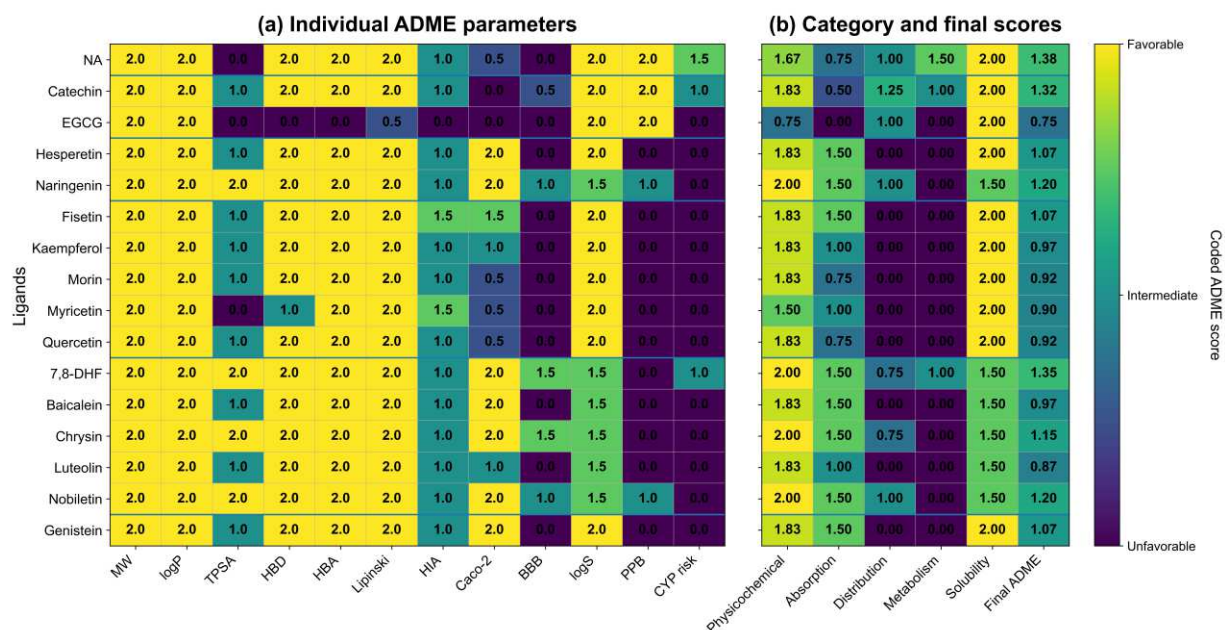


Figure 7. Consensus ADME heatmap of NA and flavonoids. (a) Individual ADME parameters converted into coded scores, where 2 indicates favorable, 1 indicates intermediate, and 0 indicates unfavorable predicted behavior. (b) Category-level scores summarizing physicochemical/drug-likeness, absorption, distribution, metabolism, and solubility properties, together with the final consensus ADME score. Final consensus ADME scores were classified using the following criteria: 1.50–2.00 = favorable, 1.00–1.49 = moderate, and 0.00–0.99 = unfavorable. Ligands are grouped according to flavonoid subclass.

NA displayed weaker TPSA-, Caco-2-, and BBB-related behavior, suggesting that its predicted pharmacokinetic profile may be limited by a highly polar surface area and reduced membrane permeability. EGCG showed the least favorable ADME profile, likely due to its high polarity, extensive hydrogen bonding capacity, and reduced permeability-related scores. This agrees with a prior review by Sahadevan et al.⁹³ that covered the application of different strategies such as structural modifications, nano-carriers as efficient drug delivery systems, synergistic studies with other bioactivities to improve the poor bioavailability, poor membrane permeability, rapid metabolic clearance and lack of stability of EGCG. Compounds such as 7,8-DHF, catechin, naringenin, and nobiletin showed moderate final consensus ADME score. Nobiletin was notable because it showed relatively favorable ADME behavior but ranked lower in the electronic descriptor analysis. Though nobiletin showed favorable drug-likeness properties, including favorable MW, logP, TPSA, hydrogen-bond donor/acceptor, and Lipinski-related scores, its final consensus ADME score was moderate because the favorable properties were offset by weaker CYP/metabolism risk-related scoring and intermediate distribution and solubility behavior. The ADME heatmap supported the interpretation that predicted ligand performance is multifactorial. Strong predicted binding to CAG-repeat nucleic acid structures may not routinely imply favorable pharmacokinetic behavior. Hence, candidate prioritization may need to integrate docking affinity,

molecular interaction patterns, DFT-derived electronic descriptors, hydration, and ADME suitability. Within this framework, compounds with favorable binding and ADME profiles may represent promising scaffolds rather than compounds that show superior performance in one specific category.

The physicochemical and solubility properties, and bioavailability score of the flavonoids and reference ligand NA was computed to evaluate predicted oral bioavailability. Table S4 of Supplementary Information lists the computed properties using SwissADME. Molecular flexibility was estimated using the number of rotatable bonds and is central in determining the oral bioavailability of small molecules. Molecules with 0–9 rotatable bonds have been proposed as favorable for oral bioavailability¹⁸. Except for EGCG, hesperetin, and nobiletin with 4, 2, and 7 rotatable bonds, respectively, all the other flavonoids have 1 rotatable bond. The reference molecule NA has 11 rotatable bonds (see Table S4). Log S (ESOL) estimates the solubility of the compounds and values in the range of –6.0 to 0.0 are considered to have favorable oral bioavailability¹⁸. All compounds had Log S (ESOL) within the –2.22 to –4.19 range, suggesting favorable oral bioavailability. SwissADME predicted a bioavailability score of 0.55 for the studied compounds except EGCG, indicating comparable and moderate predicted oral bioavailability. EGCG showed a lower bioavailability score of 0.17, consistent with its higher polarity, extensive hydrogen bonding capacity, and reduced predicted permeability. The bioavailability score alone did not strongly differentiate most ligands, and additional ADME parameters such as TPSA, Caco-2 permeability, HIA, BBB permeability, solubility, and CYP-related behavior were needed for compound prioritization.

SwissADME bioavailability radar plots shown in Figure S4 of Supplementary Information supported the ADME analysis by showing that most ligands occupied broadly acceptable physicochemical drug-like space. NA showed a relatively balanced profile, although deviations in polarity and flexibility suggest that its predicted oral drug-likeness is not ideal. Nobiletin displayed one of the most balanced profiles among the flavonoids, consistent with its favorable physicochemical and drug-likeness behavior in the ADME analysis. Several flavonols, including myricetin, quercetin, kaempferol, fisetin, and morin, showed comparable profiles, with acceptable size and lipophilicity but deviations associated with polarity and saturation. Flavones such as luteolin, baicalein, chrysin, 7,8-DHF, and nobiletin occupied a more compact physicochemical space, although individual differences were observed depending on hydroxylation and methoxylation patterns. The flavanones, hesperetin and naringenin, showed profiles reflecting greater saturation relative to flavones and flavonols, while catechin showed deviations consistent with its saturated, hydroxyl-rich scaffold. EGCG exhibited the strongest deviation from the optimal region, mainly due to excessive polarity and hydrogen-bonding capacity, consistent with its lower predicted gastrointestinal absorption and reduced bioavailability score. Additional ProTox 3.0 toxicity predictions were used as supportive in silico toxicity indicators, including predicted LD50, toxicity class, organ-toxicity alerts, carcinogenicity, mutagenicity, and CYP-related metabolism risk. Tables S5a and S5b of Supplementary Information summarize the ProTox 3.0-predicted toxicity profiles of the selected flavonoids and NA molecule, including predicted

LD50 values, toxicity classes, hepatotoxicity, carcinogenicity, neurotoxicity, mutagenicity, and CYP-related metabolism risks.

Molecular docking of flavonoids and NA with representative CAG-repeat structures

To complement the quantum mechanical descriptor analysis, molecular docking studies were performed to investigate the predicted binding affinities and binding modes of the flavonoids and the reference ligand NA with representative CAG-repeat nucleic acid structures of varying repeat lengths. Available structurally resolved CAG-repeat-containing nucleic-acid structures were obtained from the PDB, including both solution NMR and X-ray diffraction models, as summarized in Table 2. One representative structure was selected for each CAG-repeat length to enable comparison of ligand recognition across different repeat-containing nucleic acid motifs. Because CAG-repeat nucleic acid motifs can adopt multiple conformations, the selected PDB structures should be considered as representative receptor models rather than exhaustive descriptions of the full conformational ensemble. The docking analysis was interpreted at two levels: first, by comparing AutoDock Vina-predicted binding affinities across the receptors, and second, by examining whether the top-ranked poses were positioned near the noncanonical A·A mismatch motifs and supported by hydrogen-bonding, aromatic stacking, electrostatic, or other close-contact interactions.

The docking heatmap shown in Figure 8 revealed ligand- and receptor-dependent variations in predicted binding affinity across the selected CAG-repeat RNA and DNA structures. EGCG exhibited the most favorable docking scores among the studied flavonoids, with consistently favorable predicted affinities across representative CAG-repeat structures. NA showed favorable predicted binding within the CAG-repeat structures, supporting its incorporation as a reference compound for comparison (Figures 8 and 9a). Representative docking poses indicated that both NA and EGCG were positioned within groove regions near the noncanonical A·A mismatch motifs and formed intermolecular hydrogen bonds with the surrounding nucleobases and phosphate backbone (Figure 9a and b). Several flavonols, including myricetin, quercetin, fisetin, and morin, showed consistently favorable binding across the receptors, suggesting that hydroxylated, π -conjugated flavonoid scaffolds may facilitate the recognition of noncanonical A·A mismatch motifs within CAG-repeat nucleic acid structures. Receptor dependent differences were also observed, with PDB IDs 4YN6 and 9KAD generally producing more favorable predicted affinities than PDB IDs 3NJ7 and 8WU5. This suggests that inherent differences in nucleic acid conformation and local structural variations influence binding with the flavonoids and the reference ligand NA.

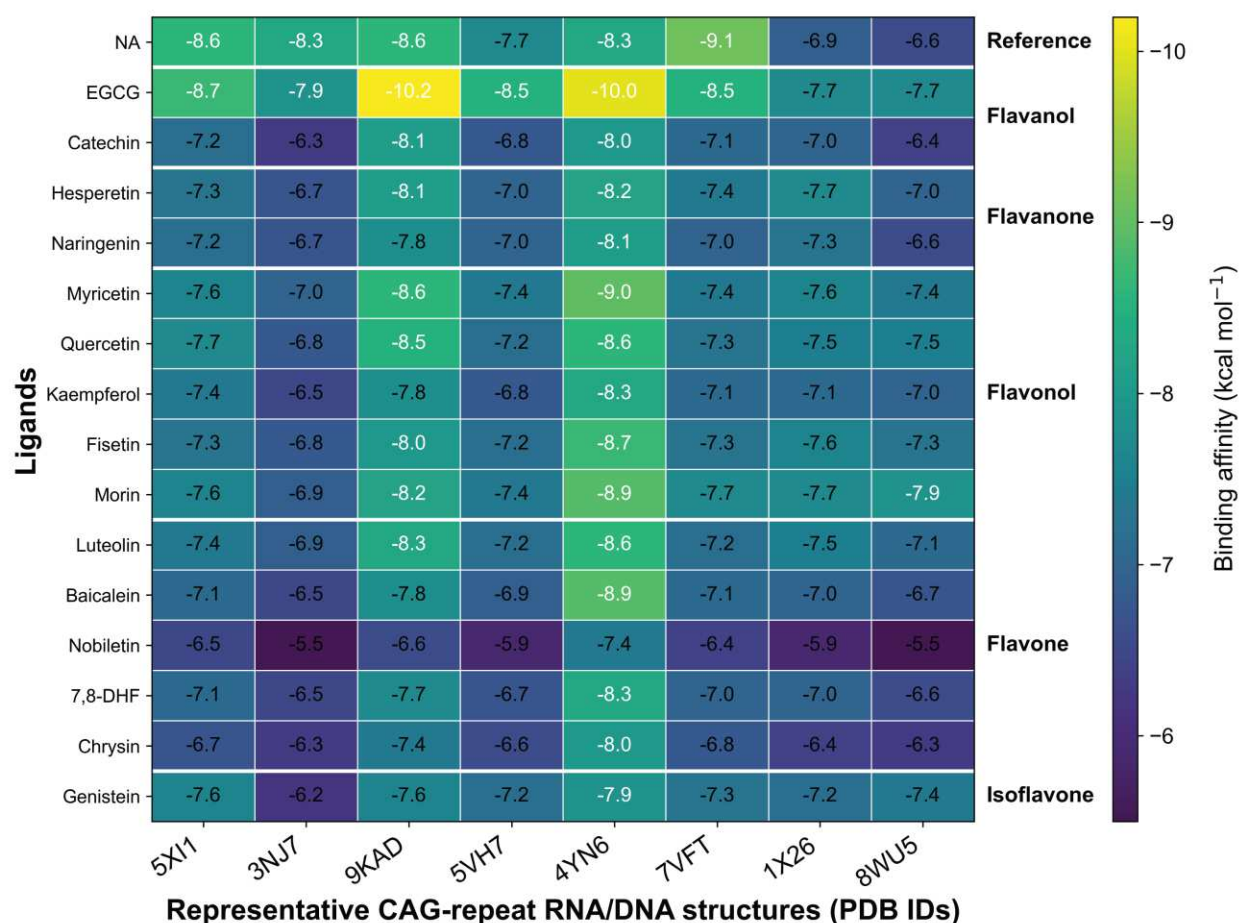


Figure 8. Heatmap of predicted binding affinities for NA and selected flavonoids docked against representative CAG-repeat RNA and DNA structures. Binding affinities are reported in kcal/mol, with more negative values indicating stronger binding. Ligands are grouped according to flavonoid subclass, and NA is the reference ligand. Grouping the ligands by subclass further highlighted trends in binding affinity across the flavonoid dataset.

Representative docking poses were then studied to evaluate whether favorable docking scores corresponded to plausible binding orientations near the A·A mismatch regions. Although EGCG was not among the most favorable candidate based on DFT-derived electronic descriptors or ADME predictions, docking results showed a more favorable predicted binding than the other flavonoids and NA (Figures 8 and 9b). EGCG binding may be driven by hydroxyl group content, hydrogen bonding, and conformational flexibility rather than by electronic softness alone, which allows EGCG to form multiple intermolecular contacts within CAG-repeat structures (Figure 9b). Flavonols such as myricetin, quercetin, and morin showed favorable docking with a model CAG-repeat structure (PDB ID: 4YN6) as shown in Figure 9c-e, likely reflecting the combined effects of π -conjugation, hydroxyl group substitution, and hydrogen-bonding. Representative docking poses suggested that myricetin, quercetin, and morin adopted intercalation-like binding orientations between adjacent nucleobase pairs rather than occupying the groove-associated regions. Nobiletin was the least favorable docking candidate (Figures 8 and 9f), with consistently

lower binding affinity values, which may be attributed to the lack of hydrogen bond donor sites in the molecule.

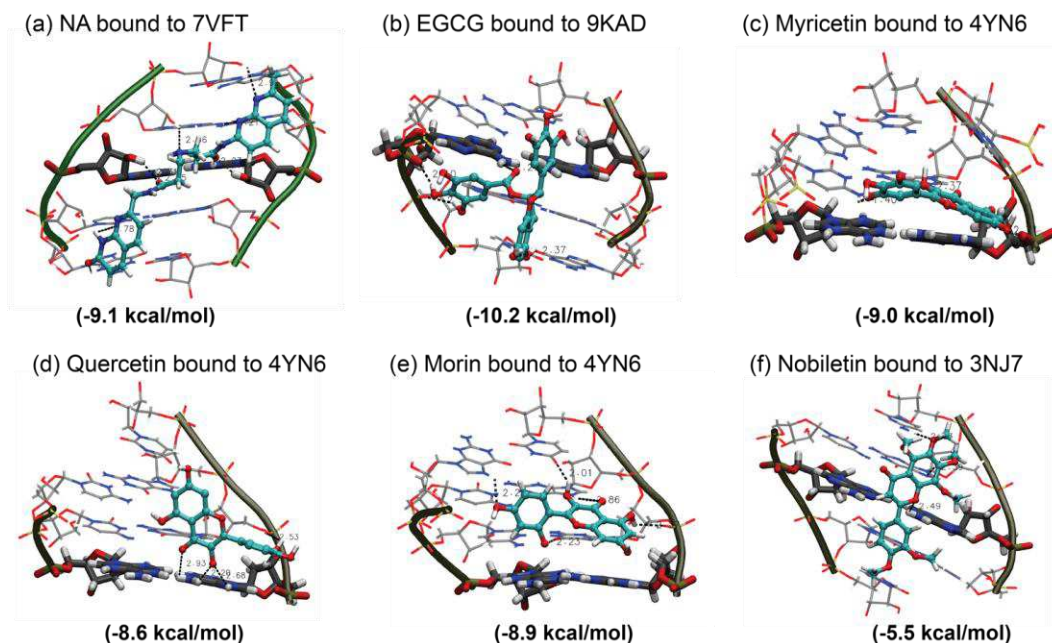


Figure 9. Representative docking poses of selected ligands bound to CAG-repeat nucleic-acid structures. NA and selected flavonoids were chosen based on their predicted docking scores and electronic descriptor profiles. Ligands and the A·A noncanonical pairs are shown in stick representation, and the intermolecular distances are labeled. Predicted hydrogen bonding, stacking, and electrostatic interactions highlight the molecular features contributing to ligand recognition of CAG-repeat motifs.

Docking and DFT descriptor analyses showed complementary trends, supporting the utilization of molecular descriptors to rationalize, but not independently predict, the observed docking behavior. Comparison of the ADME heatmap with the electronic descriptor and docking analyses indicates that ligand performance is not governed by a single property. Compounds with extensive hydroxylation and favorable solvation or binding preference, such as EGCG and selected flavonols, showed preferential predicted target-recognition behavior but less favorable ADME profiles due to increased polarity and permeability-related limitations. Nobiletin showed comparatively favorable ADME properties but weaker predicted binding, suggesting that improved pharmacokinetic balance may occur at the expense of hydrogen bonding and the extent of nucleic acid interaction. These findings highlight the need to balance electronic reactivity, hydration behavior, docking performance, and ADME predictions when prioritizing flavonoids.

Correlation analysis of molecular descriptors and docking predicted binding affinity

To further evaluate whether the calculated molecular descriptors were associated with docking predicted binding affinity, correlation analyses were performed between selected descriptors and the mean predicted binding strength as shown in Figure 10. Binding strength was calculated as the negative of the mean predicted binding affinity across the representative CAG-repeat RNA/DNA

structures, such that higher values indicate stronger predicted binding. A moderate positive correlation was observed between μ and predicted binding strength ($R = 0.62$, $p = 0.011$) as shown in Figure 10a. Because μ values are negative, the correlation plot indicates that ligands with less negative μ values generally exhibited stronger predicted binding across the CAG-repeat structures. In conceptual DFT, less negative μ values are associated with a greater tendency to release electron density and relatively higher electronic reactivity. EGCG and NA showed high predicted binding strength and relatively less negative μ values, whereas several weaker binding flavones, such as nobiletin and chrysin, showed more negative μ values. However, the relationship was not absolute; catechin showed a less negative μ value but only moderate binding strength, indicating that electronic reactivity alone does not entirely determine the docking behavior.

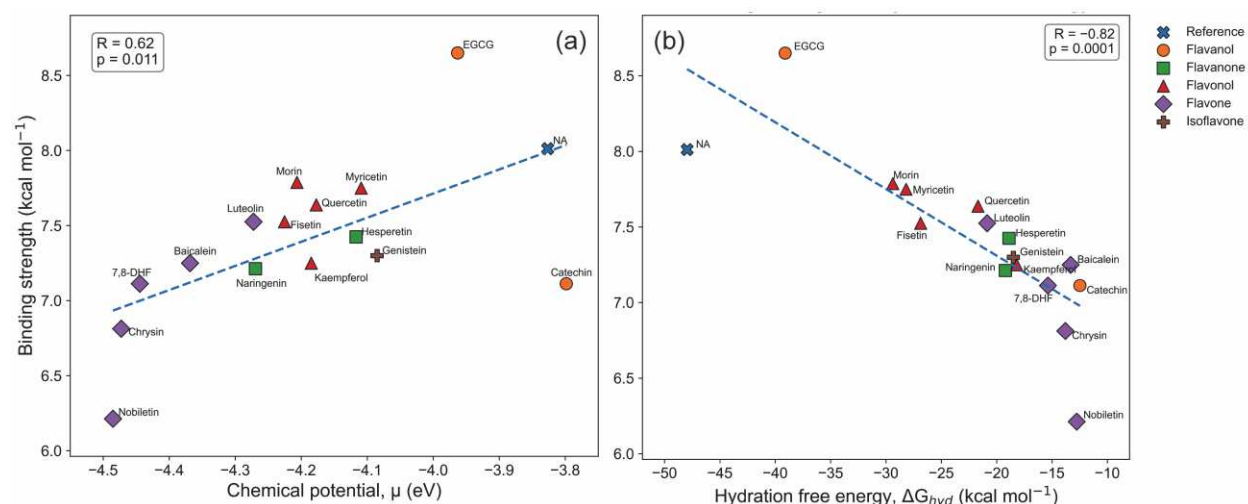


Figure 10. Correlation analysis between selected molecular descriptors and mean predicted binding strength. Binding strength was calculated as the negative of the mean predicted binding affinity across representative CAG-repeat RNA/DNA structures. A higher value indicates stronger predicted binding. Panels show correlations between binding strength and (a) chemical potential (μ) and (b) hydration free energy (ΔG_{hyd}). The dashed lines represent linear regression fits, and Pearson correlation coefficients (R) and p -values are shown in each panel.

A better correlation was observed between ΔG_{hyd} and predicted binding strength ($R = -0.82$, $p = 0.0001$) as shown in Figure 10b. The negative correlation indicates that ligands with more negative ΔG_{hyd} generally exhibited stronger predicted binding. Thus, favorable aqueous stabilization and polar interaction capacity facilitate recognition of CAG-repeat nucleic-acid structures. Strongly hydrated ligands, including NA and EGCG, followed by myricetin, morin, fisetin, and catechin, showed higher predicted binding strength than weakly hydrated ligands such as nobiletin, chrysin, baicalein, and 7,8-DHF. The correlation results suggest that predicted binding strength is influenced by both electronic and solvation-related properties. The chemical potential correlation suggests a contribution from electron donating tendency, whereas the ΔG_{hyd} correlation highlights the importance of aqueous stabilization and polar interaction capacity. Nevertheless, the

presence of outliers and the incomplete descriptor–binding relationships indicate that ligand recognition of CAG-repeat RNA/DNA motifs is multifactorial.

Integrated candidate prioritization and future experimental validation

The results reported in this study indicate that flavonoid prioritization should not be based on docking affinity alone, because predicted CAG-repeat recognition and pharmacokinetic suitability are not always aligned. EGCG showed the strongest docking-predicted binding strength relative to NA, but its less favorable ADME profile suggests that it may be more useful as a mechanistic probe of hydroxyl-rich CAG-repeat recognition than as an immediately optimized drug-like candidate. Similarly, quercetin and myricetin showed favorable docking (binding affinity) and lower global hardness values similar to the reference ligand NA, supporting their relevance as electronically favorable flavonol scaffolds. However, their relatively high polarity and hydrogen bonding may limit permeability-related ADME behavior. In contrast, 7,8-DHF, catechin, and naringenin showed more balanced predicted ADME profiles, although they were less prominent in docking or electronic descriptor analyses.

The findings suggest that future studies should evaluate two complementary directions: binding-prioritized flavonoids, such as EGCG, quercetin, myricetin, morin, and fisetin, for direct CAG-repeat binding assays, and ADME-balanced flavonoids, such as 7,8-DHF, catechin, and naringenin, as scaffolds for optimization. Nobiletin may serve as a useful low-recognition flavonoid candidate, because despite its relatively favorable predicted ADME behavior, it showed the weakest docking and less favorable electronic descriptors among the studied flavonoids. Experimental validation using biophysical binding assays, structure-based measurements like smFRET, and cellular models will be necessary to determine whether the predicted trends in molecular recognition translate into measurable CAG-repeat targeting activity and help distinguish compounds with favorable pharmacokinetic properties from those with stronger predicted CAG-repeat-recognition potential.

Conclusion

In this study, we applied an integrated computational approach to evaluate how flavonoid electronic structure, hydration behavior, docking, and predicted ADME properties relate to recognition of noncanonical CAG-repeat nucleic-acid motifs. The results showed that flavonoid recognition is not governed by a single molecular descriptor. Quercetin and myricetin displayed low energy gaps and global hardness values comparable to NA, indicating higher reactivity, whereas EGCG showed favorable predicted docking despite having high global hardness. The docking analysis identified EGCG and several flavonols, including myricetin, quercetin, fisetin, and morin, as ligands with favorable predicted interactions across representative CAG-repeat structures. Correlation analysis further suggested that hydration free energy was more strongly associated with predicted binding strength than chemical potential, underlining the importance of polarity, aqueous stabilization, and hydrogen-bonding capacity in CAG-repeat recognition. ADME analysis revealed that favorable predicted binding does not necessarily translate into favorable pharmacokinetic behavior, as illustrated by EGCG. Comparison with NA showed that selected flavonoids can reproduce different aspects of the reference ligand's predicted CAG-repeat-

recognition profile. EGCG exceeded NA in docking-predicted binding strength, whereas quercetin and myricetin showed electronic softness that was comparable to the reference ligand NA. Morin and fisetin also showed favorable docking behavior among the flavonols. On the other hand, 7,8-DHF, catechin, and naringenin were less prominent in docking or electronic descriptor analyses but showed more balanced predicted ADME behavior. These results suggest that no single flavonoid fully matches NA across all evaluated criteria; instead, different flavonoids capture different recognition- or pharmacokinetic-relevant features.

Our findings support a multifactorial model in which electronic reactivity, hydration, molecular geometry, binding-site complementarity, and ADME suitability must be considered when prioritizing flavonoid molecules. The radar plots support the consensus ADME heatmap by showing that predicted pharmacokinetic suitability depends on a balance of physicochemical properties rather than on a single descriptor. The computational results highlight that CAG-repeat recognition depends on a balance of molecular structure, electronic reactivity, hydration, and pharmacokinetic properties, providing molecular-level insight into flavonoid scaffolds relevant to trinucleotide repeat-related disorders. To our knowledge, this work represents one of the few integrated computational studies to explore naturally occurring flavonoids as small-molecule scaffolds for targeting CAG-repeat mismatch motifs associated with trinucleotide repeat expansion disorders.

Data availability

All data generated and/or analyzed are included in this article and Supplementary Information. Additional computational data and analysis files will be made available from the corresponding author upon reasonable request.

Acknowledgements

The authors gratefully acknowledge the computational support from the Chemistry Department at New Mexico Highlands University. N. Saikia acknowledges funding from the NSF EPSCoR E-CORE RII Seed Award, Award Number: 2435071. This work used Bridges2 resources at Pittsburgh Supercomputing Center (PSC) through allocations che260024p and che250036p (N. Saikia) from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by U.S. National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

Author contributions

Nabanita Saikia: conceptualization, data curation and analysis, visualization, writing – original draft, and writing – review and editing; Kiara Garcia Homs: calculations and data analysis; Bijili Navine Nanjo: calculations, data analysis, and writing – review and editing.

Competing financial interests: The authors declare no competing financial interests.

References

1. Khan, E., Biswas, S., Mishra, S.K., Mishra, R., Samanta, S., Mishra, A., Tawani, A. & Kumar, A. Rationally designed small molecules targeting toxic CAG repeat RNA that causes Huntington's disease (HD) and spinocerebellar ataxia (SCAs). *Biochimie*, **163**, 21–32 (2019).

2. Tong, H. *et al.* Huntington's Disease: Complex Pathogenesis and Therapeutic Strategies. *Int. J. Mol. Sci.* **25**, 3845 (2024).
3. Gusella, J.F., MacDonald, M.E., Ambrose, C.M. & Duyao, M.P. Molecular genetics of Huntington's disease. *Arch. Neurol.* **50**, 1157–1163 (1993).
4. Ahamad, S. & Bhat, S.A. The emerging landscape of small-molecule therapeutics for the treatment of Huntington's Disease. *J. Med. Chem.* **65**, 15993–16032 (2022).
5. Finkbeiner, S. Huntington's Disease. *Cold Spring Harb. Perspect. Biol.* **3**, a007476 (2011).
6. Malik, I., Kelley, C. P., Wang, E. T. & Todd, P. K. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell Biol.* **22**, 589–607 (2021).
7. Hu, T., Morten, M.J. & Magennis, S.W. Conformational and migrational dynamics of slipped-strand DNA three-way junctions containing trinucleotide repeats. *Nat. Commun.* **12**, 204 (2021).
8. Chatterjee, N. & Walker, G.C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* **58**, 235–263 (2017).
9. La Spada, A.R. & Taylor, J.P. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.* **11**, 247–258 (2010).
10. Ferguson, M.W., Kennedy, C.J., Palpagama, T.H., Waldvogel, H.J., Faull, R.L.M. & Kwakowsky, A. Current and possible future therapeutic options for Huntington's Disease. *J. Cent. Nerv. Syst. Dis.* **14**, 11795735221092517 (2022).
11. Jiao, Y., Yang, L., Wang, R., Song, G., Fu, J., Wang, J., Gao, N. & Wang, H. Drug delivery across the blood-brain barrier: A new strategy for the treatment of neurological diseases. *Pharmaceutics* **16**, 1611 (2024).
12. Verma, A.K., Khan, E., Bhagwat, S.R. & Kumar, A. Exploring the potential of small molecule-based therapeutic approaches for targeting trinucleotide repeat disorders. *Mol. Neurobiol.* **57**, 566–584 (2020).
13. Waring, M.J. DNA-targeting molecules as therapeutic agents. *Royal Soc. Chem.* ch. 6, 144–174 (2018).
14. Li, J. *et al.* Naphthyridine-Benzoazaquinolone: Evaluation of a Tricyclic System for the Binding to (CAG)_n Repeat DNA and RNA. *Chem. Asian J.* **11**, 1971–1981 (2016).
15. Nakamori, M.P. *et al.* A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. *Nat. Genet.* **52**, 146–159 (2020).
16. Biharee, A., Yadav, A., Jangid, K., Singh, Y., Kulkarni, S., Sawant, D.M., Kumar, P., Thareja, S. & Jain, A.K. Flavonoids as promising anticancer agents: an in silico investigation of ADMET, binding affinity by molecular docking and molecular dynamics simulations. *J. Biomol. Struct. Dyn.* **41**, 7835–7846 (2023).
17. Mohammad Ali, F.J., Zare, F., Sakhteman, A., Bahadori, S., Seradj, H. & Emami, L. Molecular docking studies, DFT, and ADMET calculations of some flavonoids and their characteristic structural features involved in inhibition of pro-inflammatory enzymes. *Nat. Prod. Res.* **39**, 5289–5299 (2025).
18. Isika, D.K. & Sadik, O.A. Selective structural derivatization of flavonoid acetamides significantly impacts their bioavailability and antioxidant properties. *Molecules* **27**, 8133 (2022).
19. Tawani, A. & Kumar, A. Structural Insight into the interaction of Flavonoids with Human Telomeric Sequence. *Sci. Rep.* **5**, 17574 (2015).

20. Salman, A., Eid, A.H., Khalaf, S.S., El-Dessouki, A.M., El-Shiekh, R.A. & Aly, S.H. Natural products proposed for the management of Huntington's disease (HD): A comprehensive review. *Naunyn. Schmiedeberg's Arch. Pharmacol.* **398**, 16817–16847 (2025).
21. Vishwas, S. *et al.* Expanding the Arsenal Against Huntington's Disease-Herbal Drugs and Their Nanoformulations. *Curr. Neuroparmacol.* **19**, 957–989 (2021).
22. Panche, A.N., Diwan, A.D. & Chandra, S.R. Flavonoids: an overview. *J. Nutr. Sci.* **5**, e47 (2016).
23. Parmenter, B.H. *et al.* High diversity of dietary flavonoid intake is associated with a lower risk of all-cause mortality and major chronic diseases. *Nat. Food* **6**, 668–680 (2025).
24. Corrente, G.A., Malacaria, L., Beneduci, A., Marino, T. & Furia, E. Quercetin and luteolin complexation with first-row transition metals in purely aqueous solutions: stoichiometry and binding site selectivity. *Dalton Trans.* **54**, 7828–7837 (2025).
25. de Souza Farias, S.A., da Costa, K.S. & Martins, J.B.L. Analysis of conformational, structural, magnetic, and electronic properties related to antioxidant activity: Revisiting flavan, anthocyanidin, flavanone, flavonol, isoflavone, flavone, and flavan-3-ol. *ACS Omega* **6**, 8908–8918 (2021).
26. Aparicio, S. A systematic computational study on flavonoids. *Int. J. Mol. Sci.* **11**, 2017–2038 (2010).
27. Juca, M.M. *et al.* Flavonoids: biological activities and therapeutic potential. *Nat. Prod. Res.* **34**, 692–705 (2020).
28. Al-Khayri, J.M., Sahana, G.R., Nagella, P., Joseph, B.V., Alessa, F.M. & Al-Mssallem, M.Q. Flavonoids as potential anti-inflammatory molecules: A review. *Molecules* **27**, 2901 (2022).
29. Maleki, S.J., Crespo, J.F. & Cabanillas, B. Anti-inflammatory effects of flavonoids. *Food Chem.* **299**, 125124 (2019).
30. Ali, D.E., Gedaily, R.A.E., Ezzat, S.M., Sawy, M.A.E., Meselhy, M.R. & Abdel-Sattar, E. In silico and in vitro anti-inflammatory study of phenolic compounds isolated from Eucalyptus maculata resin. *Sci. Rep.* **13**, 2093 (2023).
31. Farhadi, F., Khameneh, B., Iranshahi, M. & Iranshahy, M. Antibacterial activity of flavonoids and their structure-activity relationship: An update review. *Phytother. Res.* **33**, 13–40 (2019).
32. Shamsudin, N.F., Ahmed, Q.U., Mahmood, S., Ali Shah, S.A., Khatib, A., Mukhtar, S., Alsharif, M.A., Parveen, H. & Zakaria, Z.A. Antibacterial effects of flavonoids and their structure-activity relationship study: A comparative interpretation. *Molecules* **27**, 1149 (2022).
33. Zhang, H.W. *et al.* Flavonoids inhibit cell proliferation and induce apoptosis and autophagy through downregulation of PI3Kgamma mediated PI3K/AKT/mTOR/p70S6K/ULK signaling pathway in human breast cancer cells. *Sci. Rep.* **8**, 11255 (2018).
34. Mir, S.A., Dar, A., Hamid, L., Nisar, N., Malik, J.A., Ali, T. & Bader, G.N. Flavonoids as promising molecules in the cancer therapy: An insight. *Curr. Res. Pharmacol. Drug Discov.* **6**, 100167 (2023).
35. Pietta, P.G. Flavonoids as antioxidants. *J. Nat. Prod.* **63**, 1035–1042 (2000).
36. D'Amelia, V., Aversano, R., Ruggiero, A., Batelli, G., Appelhagen, I., Dinacci, C., Hill, L., Martin, C. & Carputo, D. Subfunctionalization of duplicate MYB genes in Solanum commersonii generated the cold-induced ScAN2 and the anthocyanin regulator ScAN1. *Plant Cell Environ.* **41**, 1038–1051 (2018).
37. Barreca, M.M., Alessandro, R. & Corrado, C. Effects of flavonoids on cancer, cardiovascular and neurodegenerative diseases: Role of NF-kappaB signaling pathway. *Int. J. Mol. Sci.* **24**, 9236 (2023).

38. Li, M., Qian, M., Jiang, Q., Tan, B., Yin, Y. & Han, X. Evidence of flavonoids on disease prevention. *Antioxidants (Basel)* **12**, 527 (2023).
39. Cherrak, S.A. *et al.* In vitro antioxidant versus metal ion chelating properties of flavonoids: A structure-activity investigation. *PLoS One* **11**, e0165575 (2016).
40. Mira, L. *et al.* Interactions of flavonoids with iron and copper ions: a mechanism for their antioxidant activity. *Free Radic. Res.* **36**, 1199–1208 (2002).
41. Leopoldini, M., Russo, N., Chiodo, S. & Toscano, M. Iron chelation by the powerful antioxidant flavonoid quercetin. *J. Agric. Food Chem.* **54**, 6343–6351 (2006).
42. Martins, H.F.P., Leal, J.P., Fernandez, M.T., Lopes, V.H.C. & Cordeiro, M.N.D.S. Toward the prediction of the activity of antioxidants: Experimental and theoretical study of the gas-phase acidities of flavonoids. *J. Am. Soc. Mass Spectrom.* **15**, 848–861 (2004).
43. Altunayar-Unsalan, C. & Unsalan, O. Molecular structure, antioxidant potential, and pharmacokinetic properties of plant flavonoid blumeatin and investigating its inhibition mechanism on xanthine oxidase for hyperuricemia by molecular modeling. *ACS Omega* **9**, 13284–13297 (2024).
44. Hasan, S., Khatri, N., Rahman, Z.N., Menezes, A.A., Martini, J., Shehjar, F., Mujeeb, N. & Shah, Z.A. Neuroprotective potential of flavonoids in brain disorders. *Brain Sci.* **13**, 1258 (2023).
45. Maher, P. The potential of flavonoids for the treatment of neurodegenerative diseases. *Int. J. Mol. Sci.* **20**, 3056 (2019).
46. Kopalli, S.R. *et al.* Synaptic plasticity and neuroprotection: The molecular impact of flavonoids on neurodegenerative disease progression. *Neuroscience* **569**, 161–183 (2025).
47. Cichon, N., Saluk-Bijak, J., Gorniak, L., Przyslo, L. & Bijak, M. Flavonoids as a natural enhancer of neuroplasticity-An overview of the mechanism of neurorestorative action. *Antioxidants (Basel)* **9** (2020).
48. Baptista, F.I., Henriques, A.G., Silva, A.M., Wiltfang, J. & da Cruz e Silva, O. A. Flavonoids as therapeutic compounds targeting key proteins involved in Alzheimer's disease. *ACS Chem. Neurosci.* **5**, 83–92 (2014).
49. Khan, E., Tawani, A., Mishra, S.K., Verma, A.K., Upadhyay, A., Kumar, M., Sandhir, R., Mishra, A. & Kumar, A. Myricetin reduces toxic level of CAG repeats RNA in Huntington's Disease (HD) and Spino Cerebellar Ataxia (SCAs). *ACS Chem. Biol.* **13**, 180–188 (2018).
50. Mitrasinovic, P.M. Sequence-dependent binding of flavonoids to duplex DNA. *J. Chem. Inf. Model.* **55**, 421–433 (2015).
51. Gopinath, K., Prakash, D. & Sudhandiran, G. Neuroprotective effect of naringin, a dietary flavonoid against 3-nitropropionic acid-induced neuronal apoptosis. *Neurochem. Int.* **59**, 1066–1073 (2011).
52. Louis, H., Enudi, O.C., Odey, J.O., Onyebuenyi, I.B., Igbalagh, A.T., Unimuke, T.O. & Ntui, T.N. Synthesis, characterization, DFT, and TD-DFT studies of (E)-5-((4,6-dichloro-1,3,5-triazin-2-yl)amino)-4-hydroxy-3-(phenyldiazenyl)naphthalene-2,7-diylbis(hydrogen sulfite). *SN Appl. Sci.* **3**, 712 (2021).
53. Rebas E. Role of flavonoids in protecting against neurodegenerative diseases-possible mechanisms of action. *Int. J. Mol. Sci.* **26**, 4763 (2025).
54. Tahir, M. S. *et al.* Mechanistic new insights of flavonols on neurodegenerative diseases. *Biomed. Pharmacother.* **137**, 111253 (2021).
55. Meng-Zhen, S. *et al.* Potential therapeutic use of plant flavonoids in AD and PD. *Heliyon* **8**, e11440 (2022).

56. Kim, S., et al. PubChem 2025 update. *Nucleic Acids Res.* **53**, D1516–D1525 (2025).
57. Frisch, M. J. et al. Gaussian 16, Revision B.01, Gaussian, Inc., Wallingford CT (2016) GaussView 5.0. Wallingford, E.U.A.
58. Sabry, N.M., Badry, R., Abdel-Gawad, F.K., Elhaes, H. & Ibrahim, M.A. Electronic structure, global reactivity descriptors and nonlinear optical properties of glycine interacted with ZnO, MgO and CaO for bacterial detection. *Sci. Rep.* **14**, 22801 (2024).
59. Chen, D. & Wang, H. HOMO–LUMO gaps of homogeneous polycyclic aromatic hydrocarbon clusters. *J. Phys. Chem. C* **123**, 27785–27793 (2019).
60. Liakos, D.G., Guo, Y. & Neese, F. Comprehensive benchmark results for the domain based local pair natural orbital coupled cluster method (DLPNO-CCSD(T)) for closed- and open-shell systems. *J. Phys. Chem. A* **124**, 90–100 (2020).
61. Nagy, P. R. State-of-the-art local correlation methods enable affordable gold standard quantum chemistry for up to hundreds of atoms. *Chem. Sci.* **15**, 14556–14584 (2024).
62. Garcia-Ratés, M., Becker, U. & Neese, F. Implicit solvation in domain based pair natural orbital coupled cluster (DLPNO-CCSD) theory. *J. Comput. Chem.* **42**, 1959–1973 (2021).
63. Neese, F., Wennmohs, F., Becker, U. & Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **152**, 224108 (2020).
64. Neese, F. The ORCA program system. Wiley Interdisciplinary Reviews: *Comput. Mol. Sci.* **2**, 73–78 (2012).
65. Adu-Effah, N. & Saikia, N. Intermolecular hydrogen bonding delineates the stability of non-canonical adenine base pairs: a first-principles study. *Phys. Chem. Chem. Phys.* **26**, 29150–29166 (2024).
66. Marenich, A.V., Cramer, C.J. & Truhlar, D.G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B.* **113**, 6378–6396 (2009).
67. de Souza, J.C.G., de Souza, L.D.S.S., Hidalgo, A.F., de Azevedo, G.T., de Souza, G.L., Lacerda, C.D., Junior, S.D. & Pereira, A.M. DPPH-based antioxidant screening and cellulase-assisted hydrodistillation of matourea azurea: Improved essential oil yield and density functional theory insights into a rare terpene. *Chem. Biodivers.* **22**, e01109 (2025).
68. Saikia, N. & Deka, R.C. Theoretical study on pyrazinamide adsorption onto covalently functionalized (5,5) metallic single-walled carbon nanotube. *Chem. Phys. Lett.* **500**, 65–70 (2010).
69. Chattaraj, P.K., Sarkar, U. & Roy, D.R. Electrophilicity index. *Chem. Rev.* **106**, 2065–2091 (2006).
70. Oyewole, R.O., Oyebamiji, A.K. & Semire, B. Theoretical calculations of molecular descriptors for anticancer activities of 1, 2, 3-triazole-pyrimidine derivatives against gastric cancer cell line (MGC-803): DFT, QSAR and docking approaches. *Heliyon* **6**, e03926 (2020).
71. Bultinck, P., Langenaeker, W., Carbó-Dorca, R. & Tollenaere, J.P. Fast calculation of quantum chemical molecular descriptors from the electronegativity equalization method. *J. Chem. Inf. Comput. Sci.* **43**, 422–428 (2003).
72. Pearson, R.G. Absolute electronegativity and hardness correlated with molecular orbital theory. *Proc. Natl. Acad. Sci. USA.* **83**, 8440–8441 (1986).
73. Babu, K.J. & Ayodhya, D. Comprehensive investigation of Co (II), Ni (II) and Cu (II) complexes derived from a novel Schiff base: Synthesis, characterization, DNA interactions, ADME profiling, molecular docking, and in-vitro biological evaluation. *Results Chem.* **6**, 101110 (2023).

74. Guan, L., Yang, H., Cai, Y., Sun, L., Di, P., Li, W., Liu, G. & Tang, Y. ADMET-score - a comprehensive scoring function for evaluation of chemical drug-likeness. *Med. Chem. Commun.* **10**, 148–157 (2019).
75. Daina, A., Michielin, O. & Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **7**, 42717 (2017).
76. Dong, J. *et al.* ADMETlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *J. Cheminform.* **10**, 29 (2018).
77. Gu, Y. *et al.* admetSAR3.0: a comprehensive platform for exploration, prediction and optimization of chemical ADMET properties. *Nucleic Acids Res.* **52**, W432–W438 (2024).
78. Ding, X., Hu, X., Chen, Y., Xie, J., Ying, M., Wang, Y., & Yu, Q. Differentiated Caco-2 cell models in food-intestine interaction study: Current applications and future trends. *Trends Food Sci. & Technol.* **107**, 455–465 (2021).
79. Alnajeebi, A.M. Molecular docking, DFT analysis, pharmacokinetic profiling and MD simulation of *Ilex aquifolium* L. flavonoids as potential alpha-glucosidase predicted inhibitors for diabetes. *Sci. Rep.* **16**, 22168 (2026).
80. Pettersen, E.F. *et al.* UCSF Chimera--A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
81. Trott, O. & Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
82. Sandler, I., Chen, J., Taylor, M., Sharma, S. & Ho, J. Accuracy of DLPNO-CCSD(T): Effect of basis set and system size. *J. Phys. Chem. A* **125**, 1553–1563 (2021).
83. Ronsisvalle, S., Panarello, F., Longhitano, G., Siciliano, E. A., Montenegro, L., & Panico, A. Natural flavones and flavonols: Relationships among antioxidant activity, glycation, and metalloproteinase inhibition. *Cosmetics* **7**, 71 (2020).
84. Trouillas, P., Marsal, P., Siri, D., Lazzaroni, R. & Duroux, J.-Luc. A DFT study of the reactivity of OH groups in quercetin and taxifolin antioxidants: The specificity of the 3-OH site. *Food Chem.* **97**, 4, 679–688 (2006).
85. Brovarets, O.O. & Hovorun, D.M. Conformational diversity of the quercetin molecule: a quantum-chemical view. *J. Biomol. Struct. Dyn.* **38**, 2817–2836 (2020).
86. Choudhary, V., Bhatt, A., Dash, D. & Sharma, N. DFT calculations on molecular structures, HOMO-LUMO study, reactivity descriptors and spectral analyses of newly synthesized diorganotin(IV) 2-chloridophenylacetohydroxamate complexes. *J. Comput. Chem.* **40**, 2354–2363 (2019).
87. Hasan, A.W. *et al.* Synthesis, characterization, DFT calculation, and biological activity of a new Schiff base ligand and its ZnO and Co₃O₄ nano-metal oxide complexes. *BMC Chem.* **20**, 10 (2026).
88. Abdulazeez, I., Khaled, M. & Al-Saadi, A.A. Impact of electron-withdrawing and electron-donating substituents on the corrosion inhibitive properties of benzimidazole derivatives: a quantum chemical study, *J. Mol. Struct.* **1196**, 348–355 (2019).
89. Puthanveedu, V., Pulikkool, C., Poonkottil, N. & Muraleedharan, K. Theoretical probing to the reactivity and biological effects of the phytochemical, coumestrol and its derivatives. *Chem. Phys. Impact* **4**, 11 (2022).

90. Juma, J.M. & Vuai, S.A.H. Computational studies of the thermodynamic properties, and global and reactivity descriptors of fluorescein dye derivatives in acetonitrile using density functional theory. *J. Chem. Res.* **45**, 800–805 (2021).
91. Zafar, A. & Reynisson, J. Hydration free energy as a molecular descriptor in drug design: A feasibility study. *Mol. Inform.* **35**, 207–214 (2016).
92. Fossat, M.J., Zeng, X. & Pappu, R.V. Uncovering differences in hydration free energies and structures for model compound mimics of charged side chains of amino acids. *J. Phys. Chem. B* **125**, 4148–4161 (2021).
93. Sahadevan, R., Singh, S., Binoy, A. & Sadhukhan, S. Chemico-biological aspects of (-)-epigallocatechin-3-gallate (EGCG) to improve its stability, bioavailability and membrane permeability: Current status and future prospects. *Crit. Rev. Food Sci. Nutr.* **63**, 10382–10411 (2023).

Supplementary Files

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

- [SupplementaryInformation.docx](#)