Antioxidant properties of phenolic compounds

Background:
Resulting from an excess of reactive oxygen species (ROS), states of oxidative stress are known to be causative factors in a number of human illnesses. Antioxidants are substances that are easily oxidized. Antioxidants react quickly to scavenge the free radicals generated by ROS; this lessens the chemical damage associated with free radical induced oxidation of proteins, nucleic acid, and lipids.

Phenolic compounds are well known antioxidants; many of the nonalcohol-related health benefits of wine consumption have been attributed to the presence of polyphenolic substances. It is generally considered that the primary mechanism of the radical scavenging activity of polyphenols is hydrogen atom donation. The corresponding antioxidant radical formed (from hydrogen donation) is either unreactive, often as a result of steric hindrances, or undergoes a subsequent electron rearrangement to a more stable form.

Quantitative Structure Activity Relationships (QSAR) represent a widely used approach to predict chemical or biological properties using a set of compound-specific set of variables that often include chemical/physical properties or quantum-chemical computed parameters. Previously published studies have shown that phenolic oxidation potentials (E) are closely correlated with experimentally observed antioxidant properties.

Objective:
This project involves the development and validation of a QSAR model to predict antioxidant properties of phenolic substances. Using reported oxidation potentials for a range of phenolic structures, a data set of properties for each of these structures has been generated. These data have been divided into two subsets: one that will be used to develop a mathematical model to predict oxidative potentials and a second smaller subset that will be used to test the model's effectiveness. Based upon the model developed, new phenolic structures having improved antioxidant properties will be explored and evaluated with the developed QSAR model.

Data Matrix Development: The first step in QSAR is to develop a data matrix of sample compounds and property variables. By convention, each matrix row corresponds to a different sample (compound) and each data matrix column corresponds to a different variable (property or computed molecular parameter). QSAR models are often designed for prediction of a biological effect such as enzyme inhibition or toxicity (QSAR has been particularly effective for lessening the degree and the extent of required animal toxicity testing).

Biological Effect (dependent variable): For this project, the predicted (dependent) variable to evaluate biological effects for phenolic compounds will be oxidation potentials; these potentials are closely correlated with experimental antioxidant properties. ***Table 1*** presents oxidation potential data that have been reported for 41 phenolic compounds having various substituent groups on a phenol backbone (***Figure 1***). The oxidation potential, E_7, is an indication of the voltage required to remove an electron from the oxygen atom present within the phenolic backbone.

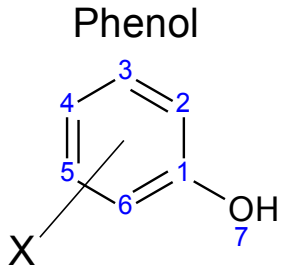

Figure 1: *Phenol backbone structure being used for this study*

Compound	***X Substituent***	***E_7 (Volts)***	***Compound***	***X Substituent***	***E_7 (Volts)***
1	4-NO_2	1.23	22	3-OH	0.81
2	3-Cl, 5-Cl	1.15	23	2-OCH_3	0.77
3	4-CF_3	1.13	24	4-OCH_3	0.73
4	3-NO_2	1.13	25	3-OCH_3, 4-OCH_3	0.67
5	4-COPh	1.12	26	3-OCH_3, 4-OCH_3, 5-OCH_3	0.66
6	3-CN	1.11	27	Sesamol	0.62
7	4-COOH	1.04	28	2-OH, 4-COOH	0.6
8	3-$COCH_3$	0.98	29	2-OCH_3, 6-OCH_3	0.58
9	4-H	0.97	30	2-OH, 3-OH	0.58
10	4-Br	0.96	31	2-OH,3-OH, 5-$COOCH_3$	0.56
11	4-Cl	0.94	32	3,4-Dihydrocynnamic acid	0.54
12	4-F	0.93	33	2-OH	0.53
13	Tyrosine	0.89	34	2-OH, 4-CH_3	0.52
14	3-OH, 4-$COCH_3$	0.89	35	4-OH	0.46
15	4-CH_3	0.87	36	4-NH_2	0.41
16	3-OCH_3, 5-OCH_3	0.85	37	4-CN	1.17
17	3-CH_3	0.85	38	4-$COCH_3$	1.06
18	3-OH, 5-OCH_3	0.84	39	4-t-Butyl	0.8
19	3-CH_3, 5-CH_3	0.84	40	2-CH_3, 6-CH_3	0.77
20	4-Phenyl	0.84	41	2-OCH_3, 4-CH_3	0.68
21	2-CH_3	0.82			

Table 1: *Table of oxidation potentials for 41 phenolic compounds having various X substituents on a phenol backbone (**Figure 1**).*

QSAR Independent Variables (physical/chemical properties, computed quantum mechanical parameters)
This project incorporates a number of **computated** molecular properties—determined using either Chemsketch or Spartan software--to explore as possible variables to use for predicting oxidation potentials. These include: Log P, Polarizability, Surface Tension, Index of Refraction, Density, Vertical (Koopman's) Ionization energy ($-E_{HOMO}$), Charge on the O_7 atom, O_7-H vibrational frequency, Energy of LUMO-r (of radical form), Electron Affinity, Vertical Ionization Potential, Molecular Area, Molecular Volume, Dipole Moment, Electronegativity, Hardness, Softness, and Electrophilic Index.

QSAR Model Development Requirement

Overall Goal: Develop and evaluate a three-variable mathematical model to predict phenolic antioxidant activity (quantified by the oxidative potential) in the form:

Oxidative potential = a * variable1 + b* variable2 + c * variable3 + d

- ***Data Set***: You are being provided a data set for the 41 compounds with their respective oxidative potential along with 18 other properties/variables—5 determined using ChemSketch, 13 obtained from Spartan quantum chemistry methods using Spartan software or calculated from Spartan results.

- ***Requirements***:

    - Use these data to generate a correlation matrix (***Tools, Data Analysis, Correlation***) that shows the correlation coefficient for all pairs of these 19 parameters. The correlation coefficients vary between -1 to +1; an r value of 0 indicates no correlation and thus little utility in predicting E….

    - Identify the variable with the best correlation to the oxidative potential E. Conduct a linear regression (***Tools, Data Analysis, Regression)*** to predict E using this independent variable. Identify the Coefficient of Determination R^2 that represents the fraction of the overall variability accounted for by this single variable model.

    - Identify other variables that work well with the initially selected variable; conduct a series of exploratory regressions using two independent variables (one being the first selected) and find which pair gives the highest Coefficient of Determination R^2 for predicting E. The best two pairs will provide the 2nd and 3rd best variable to use in your final model.

    - Conduct a three independent variable model using the selected variables to predict E.

- ***Documentation***:
    - Write an abstract giving a brief introduction, outlining what you did, and summarizing the results you obtained. The results should include an equation for each of your three models, the corresponding coefficients of determinations, and the respective root-mean-square errors from both the calibration and test sets.
    - Provide a spreadsheet that clearly shows:
        - A table showing the correlation matrix that you determined.
        - Three plots of experimental vs predicted values for each of your three models.