EVANS (EigenValue Analysis) Manual

1. Introduction

1.1 Overview

EVANS is a chirality-sensitive 3D QSPR methodology designed for predicting important attributes of potential drug molecules. This methodology offers several advantages over existing QSAR/QSPR methods, including the elimination of an alignment step, chirality considerations, and unbiased predictions. EVANS has been extensively tested across various pharmacological endpoints using diverse datasets and chemometric methods. Importantly, EVANS is open source.

1.2 Key Features

- Chirality-sensitive 3D QSPR methodology
- No alignment step required
- Broad applicability
- Open-source

2. Publications

For a detailed understanding of EVANS, please refer to the following publications:

1. **Eigen Value ANalySis (EVANS) – A Tool to Address Pharmacodynamic, Pharmacokinetic and Toxicity Issues**
    - Joseph, B., Gomatam, A.N., Shaikh, M.A.S., Khedkar, V., Martis, E.A.F., Coutinho, E.C., 2019.
    - Read Article
2. **How effective are ionization state-based QSPKR models at predicting pharmacokinetic parameters in humans?**
    - Gomatam, A., Joseph, B., Advani, P., Shaikh, M., Iyer, K., Coutinho, E., 2022a.
    - Read Article
3. **Predicting toxicity of endocrine disruptors and blood-brain barrier permeability using chirality-sensitive descriptors and machine learning**
    - Gomatam, A., Joseph, B., Gawde, U., Raikuvar, K., Coutinho, E., 2022b.
    - Read Article

3. Code and Steps on GitHub

The code for EVANS is available on GitHub. Below is a description of the steps and code for implementing EVANS.

3.1. Molecular Simulations

The first step in EVANS involves a short molecular simulation of the ligand molecules to obtain stable starting structures. For reference, we provide the protocol and scripts for running simulations in the AMBER molecular dynamics program.

- **Scripts:**
    1. `no_conect.sh`: Removes CONECT records from input structures (sdf or mol2)
    2. `auto-antechamber.sh`: Ligand parametrization using the Antechamber package
    3. `autoleap.sh`: Creates topologies and initial coordinates in AMBER format using the LEap program
    4. `runMin.sh`: Runs equilibration MD simulation protocol using input files (min.in for minimization, heat.in for heating, density.in for density equilibration, and prod.in for production MD)
    5. `remin.sh`: Trajectory analysis and reminimization of the last MD frame

3.2. Distance Calculation

Interatomic distances (measured as the distance between the centroid of each molecule, and the centroid of each atom pair in the molecule) are calculated using the Visual Molecular Dynamics program.

- **Script:**
    - `dist_CentBond-CentMol.tcl` VMD script for computing molecular interatomic distances between each atom pair and the centroid of the molecules.

3.3. Matrix Generation and Processing

Interatomic distances are plotted on the off-diagonal elements of a 2D matrix. For chiral compounds, distances are calculated separately for each enantiomer, and the maximum and minimum distances are plotted. The diagonal elements are populated with physicochemical properties computed using the PaDEL program (or any freely available tool). The physicochemical property integrated distance matrix is diagonalized to generate covariance matrices.

- **Script:**
  - `Evans_v2020.1.xlsm:` Excel macro for matrix generation and processing.
- **Tasks include:**
  - Canonical numbering using the Morgan algorithm
  - Distance matrix generation
  - Physicochemical property-integrated distance matrix generation
  - Diagonalization to generate covariance matrices
- **Note:** Physicochemical properties can be calculated using any open-source tool.

3.4. Eigenvalue Calculation

Covariance matrices are mathematically transformed to their corresponding eigenvalues, which are used as descriptors for QSAR/QSPR model building.

- **Script:**
  - `newCoVarMats_2_eigen_tab.r:` R script for computing eigenvalues from covariance matrices

3.5. ML Model Building

The eigenvalue descriptors are correlated with the corresponding biological property using suitable chemometric methods. These can be implemented using any machine learning workflow. Sample python ML workflows are provided as shown below.

- **Scripts:**
  - `classification.py:` Python script for building machine learning models (classification)
  - `regression.py:` Python script for building machine learning models (regression)