```
import pandas as pd
import math as m
import numpy as np
import scipy.stats as stats
import time
import csv

from rdkit import Chem
from rdkit.Chem import Descriptors



def main():
    start_time = time.clock()

    # Read in train and test as Pandas DataFrames
    df_train_all = pd.read_csv('train.csv',usecols=["smiles"])
    # df_test_all = pd.read_csv('test.csv',usecols=["smiles"])

    # df_all = pd.concat([df_train_all,df_test_all])
    functions = [i for i in dir(Descriptors) if not (i.startswith("__") or i

    with open("train_rdkit_new.csv",'wb') as w:
        writer = csv.writer(w)
        writer.writerow(list(['smiles'])+list(functions))
        for i, row in df_train_all.iterrows():
            holder = list([row.smiles])
            mol = Chem.MolFromSmiles(row.smiles)
            print "%sth row is processing"%i
            # print mol.GetNumAtoms()
            for each in functions:
                item = getattr(Descriptors,each)
                if callable(item):
                    try:
                        holder.extend([item(mol)])
                    except AttributeError:
                        print "AttributeError"
                    except ValueError:
                        print "ValueError"
            writer.writerow(holder)

    print "Runtime is ", time.clock() - start_time, "seconds"
```

```
47 if __name__ == "__main__":
48     # execute only if run as a script
49     main()
50
```