Subject: Re: drug score/evolutionary algorithm
Posted by pc419714@ohio.edu on Fri, 05 Apr 2019 00:56:06 GMT
View Forum Message <> Reply to Message

Please correct me if I am wrong--- For our evolutionary algorithm we selected 400 generations, 4096 compounds per generation, and 32 surviving each generation and set it so that it generates structures like approved drugs. The fitness was set to skelsphere similarity with a weight of 1.

After looking at the source code I have come to the following conclusions. Data Warrior uses java's chemistry development kit. Each generation, it finds the compounds most similar to the parent compound based on skelsphere similarity while leaving the selected structure untouched using fragments from known drugs. It selects the 32 most similar to the parent each generation as the most fit because that's what we set it to. It's not calculating fitness based on the drug score. The drug score is only calculated when you select calculate properties after generating all the compounds.

Possible mutations-- add atom, insert atom, change atom, cut out atom, delete atom, add bond, change bond, delete bond, change ring, group migration, swap substituent, delete substituent, cutout fragment. After generating a mutation, it checks to see if the structure has proper valence and if there is ring strain. It also calculates the frequency of the mutation in the population to determine the probability of each mutation. The probability would really depend on your specific molecule and if what's being generated is a valid molecule.

The drug score was that equation I sent you in the previous email-- it's not used in the evolution. It's generated using that equation I sent you in the previous email. Drug Score is different from druglikeness. Druglikeness= nasty incrementsum + increment sum / sqrt(fragment count)

drug
score=(0.5+0.5/(1+exp(cLogP-5))*(1-0.5/(1+exp(cLogS+5))*(0.5+0.5/(1+exp(0.012*Molweight-6))*(1-0.5/(1+exp(Druglikeness))*if(Mutagenic=="high",0.6,if(Mutagenic=="low",0.8,1))*if(Tumorigenic== "high",0.6,if(Tumorigenic=="low",0.8,1))*if(ReproductiveEffective== "high",0.6,if(ReproductiveEffective=="low",0.8,1))*if(Irritant== "high",0.6,if(Irritant=="low",0.8,1)

Data warrior assigns nasty functional groups various scores called increments. It also gives compounds with more than 50 druglike fragments as more druglike.

The toxicity predictor relies on compounds from the RTECS database. I can probably request it from the database site.

Thanks so much. They just want to know as many details as possible. I've been trying to write my own machine learning algorithms in rdkit. I haven't played much with chemistry development kit.

Patrick