Subject: non-toxic Build Evolutionary Library
Posted by juliocoll on Wed, 28 Jun 2023 08:18:45 GMT
View Forum Message <> Reply to Message

dear Thomas,

I wonder if a check by non-toxic criteria could be included into the "build evolutionary library" right after each new molecule is being generated and before selection to be included in the final table.

That would increase the numbers of useful molecules at the final table and avoid waste of resources like memory during the evolutionary proccesses.

The possibility was raised because in one extreme case, I generated ~6000 BEL molecules, but only ~200 (~3 %) were non-toxic. As examined, most of the toxic BEL contained a "tertiary immoniun" nasty function (see part of the file attached)

Thank you for your attention, sincerely
julio

File Attachments
1) 2427v2Evolutionary_Library.dwar, downloaded 493 times

Subject: Re: non-toxic Build Evolutionary Library
Posted by juliocoll on Wed, 23 Aug 2023 07:52:42 GMT
View Forum Message <> Reply to Message

Dear Thomas
Thanks to Norwith B.

, a few days ago I detected a "prefer toxic risk" criteria added to the build evolutionary library:
THANKS A LOT!!!

I am experimenting now with a coouple of models using simultaneous docking score, mol weight,logP, and presence or absence of toxic risk criteria with different thresholds and weights.

I will comment the results on due time to the forum

Cheers
Julio

Subject: Re: non-toxic Build Evolutionary Library
Posted by juliocoll on Mon, 28 Aug 2023 10:12:31 GMT
View Forum Message <> Reply to Message

Dear Thomas,
Thanks to Norwid awardeness of the toxicity risk criteria option to the old DW updates at the end of the "Build Evolutionary Library"(EL) Add Criterion, I am experimenting with different protein/ligand pairs by adding it to my usual DW docking score, molecular weight and logP criteria that I called "evolutionary dockings".
I focus my comments now to preliminary results predicted with one model previously described for the ratVKORC1 protein/brodifacoum ligand complex (Coll, 2023 vs2 https://chemrxiv.org/engage/chemrxiv/article-details/64c8ddd 1ce23211b20ee37e3).

1) PRELIMINARY EXPERIMENTS
I first made some preliminary experiments with different ± toxicity risk values. The EL version was the most recent Windows update of 10/08/2023 which included the toxicity risk in an e9 of 47 CPU provided with 64Gb of RAM. Additional criteria included, DW docking score to the VKORC1/brodifacoum complex (Coll, 2023vs2) weight x4, molecular weight <400 x2, logP <3 x1 and ± toxicity risk values x4. Other running variables were: cycle: automatic, total run count=1, compounds per cycle=128, compounds survive a cycle =16 , Create compounds like=Approved drugs. Results in Table and graphic:

1) Numbers of raw children. There were similar numbers of randomly generated raw children (13033-15827) for toxic risks 0, 0.3, 1 and 2. The lowest numbers were those obtained at the highest toxic risk=4. Maximal fitting to the set criteria were very similar for all cases (>0.89). Maximal docking affinities of the nontoxi-macro children were all higher than 114 DW docking-score unitless. Norwid Behrnd discovered that the number of raw children divided by the number of cycles was nearly constant ~100 (now confirmed in 6 other protein/ligand models). Norwid sugested that "there seems to be a compensation by DW, it keeps generating new molecules until enough spots per generation are allocated". It remains unclear whether or how that is accomplished by EL.

2) Numbers of fitted children. The numbers of raw children fitting the targeted criteria were between 1393 and 2691, except for those that were lower at the maximal toxic risk=4 (977).

3) Numbers after nontoxi-macro. The percentages of children fitting all the criteria which remained toxic (nontoxi-macro) as detected by the NorwidJulio's macro after EL, were minimal at risks 0.01, 0.1, 0.3, and 1 (100% may be expected if the criteria were not a preference). It looks like the risk option=0 doesn’t work as expected. Is that because is a prefernce or because trouble interpreting the "<"=0 risk?. Surprisingly, the toxic risk option =4 did not generated any nontoxic-macro children (?).

4) Main conclusions: With the above mentioned model and criteria, the toxic risk value =1 generated the highest number of nontoxi-macro children (2330). An ~ 2.6 fold higher than that obtained without using any toxic risk criteria (893). The recently introduced toxicity risk criteria seem to perform as expected during individual molecular docking evolution (confirmed in 6 differn protein/ligand published pairs). The final/manual Norwith/julio's macro serves to detect any toxic "survivors" that may remain after the EL with toxicity risk criteria. Their residual presence may be explained because the toxicity risk criteria is a preference rather than a threshold.
These are preliminary results because i) only one protein/ligand model was analysed, b) stochastic variations may be expected when more than 1 run would be applied, and c) all the EL criteria are defined at DW as preferences rather than as thresholds.

File Attachments
1) Thomas Tables Agosto tocxicityRisks.docx, downloaded 416 times

Subject: Re: non-toxic Build Evolutionary Library
Posted by juliocoll on Mon, 28 Aug 2023 10:25:36 GMT
View Forum Message <> Reply to Message

PRELIMINARY REMARKS
- The toxicity risk added by Thomas to work during EL is a new important feature that allows ~2-fold increase in the number of molecules explored for best docking during 3 consecutive runs, taking only 6-24h per pair to complete with one appropriated high-RAM computer.

- The resulting children still required a macro filter at the end of the runs to fine tune the elimination of toxicities "surviving" the preferenced toxicity risk criteria

- Any further improvements in the toxicity risk will most probably be reflected on further yield improvements in non-toxic children

To best understand the EL selective process for fitnes, some questions still remain since it is unclear how weight calculations decide which raw children may be included into the fitting children list. For instance:

- Are the final toxicity risks altered by the rest of the criteria weights?.
- Do each of the weight criteria act independently?.
- Are toxicity, nasty functions and risks applied as filters to all molecules in the same cycle or one by one after been generated?.
- Would a mathematical formulation help to clarify these possibilities?.
- Is there any room for further improvement?

Thanks Thomas for your patience!!

Subject: Re: non-toxic Build Evolutionary Library
Posted by thomas on Fri, 08 Sep 2023 10:37:21 GMT
View Forum Message <> Reply to Message

I just realize that while fuzzy score and fitness calculation are very similar and partially use the same default values, the function for calculating the score are different. Also in case of fitness it is not obvious how the score is calculates from the real values. Probably, I should use the fuzzy score UI also for the fitness panels. But for the time being, the calculation for the property fitness scores is as follows:

- If the property value is between the desired min and max, then the fitness score is 1.0. If the property is above max or below min, then the score is calculated as score=e^(delta/const) where delta is max-property or property-min, respectively, and const is a property dependent constant that defines how quickly the score decreases from 1.0 when the property departs from the allowed window. Currently, for cLogP const=0.5 and for toxicity const=3.0. This causes a cLogP score 0f 1/e if a cLogP value is 0.5 off the window and a tox score of 1/e if the toxicity indication sum is 5 (assuming that 2 is the allowed upper limit. My current gut feeling says that const should be higher for cLogP and lower for tox risk. In any rate it would be better, if this could be changed from a default value in the fitness panels by the user.

As a reminder: toxicity indication sum is the sum of toxity indation values over all four toxicity classes where a low risk label contributes a 1 and a high risk label contributes a 2. Thus, the maximum would be 8 for a compound that is classified 'high risk' in all four categories.

Once we have the fitness values of all individual properties (0.0 to 1.0), the overall fitness is calculated as a geometric mean of all contributing values considering the defined contribution weight. This is done by calculating the product of all weight*fitness values and then drawing the n-th root from it, where n is the sum of all weight values.

This should answer all the questions I hope.

Subject: Re: non-toxic Build Evolutionary Library
Posted by juliocoll on Mon, 11 Sep 2023 17:02:38 GMT
View Forum Message <> Reply to Message

Thank you Thomas.

Although quite complex, it is becoming clearer. Little by little.....

Subject: Re: non-toxic Build Evolutionary Library
Posted by juliocoll on Sun, 08 Oct 2023 20:50:27 GMT
View Forum Message <> Reply to Message

In case any of you would be interested in finding more about possible docking applications for the Build Evolutionary Library, BEL of recent versions of Data Warrior,

BEL co-evolution criteria was comparatively applied as a preliminary attempt to improve putative control predictions of coagulation, antibiotics, cancer, diabetes, or fibrosis.

The report, included 9 protein targets (Vkorc1, FtsZ, Sglt1, Vegfr, Nkcc1, Hsp47, Sglt2, P2x3r, LolEC) ,their corresponding reference ligands (parents) and their corresponding dwar and pdb data files.
It can be found at:

https://chemrxiv.org/engage/chemrxiv/article-details/6512b16 2ade1178b2424c325

More fine-tunings still being cooked.....