

# Computer Modeling of Biomolecules



Final Project

Νίκος Περδικοπάνης  
nikosp@di.uoa.gr

Ιανουάριος 2015

## Project: Re-ranking Virtual Screening results

- Στόχος είναι η εύρεση inhibitors (screen out) οι οποίοι είναι επιλεκτικοί για την πρωτεΐνη ενδιαφέροντος.
- Τη είσοδο αποτελούν τα αποτελέσματα (virtual screening results ) των midterm projects (ranked compounds in sdf format)

# Ανάπτυξη λογισμικού-Πλατφόρμα

- Για την υλοποίηση χρησιμοποιήθηκαν:
  - Python scripting (2.7.10)
  - Platform : Ubuntu with Apache web server
- Αναπτύχθηκαν
  - Command line tools για τα επιμέρους τμήματα της εφαρμογής
  - Web Interface για την απομακρυσμένη πρόσβαση και εκτέλεση.

# Μέρος Πρώτο Command Line

- createTabFileCLI.py

>python createTabFileCLI.py filename1 ....filenameN

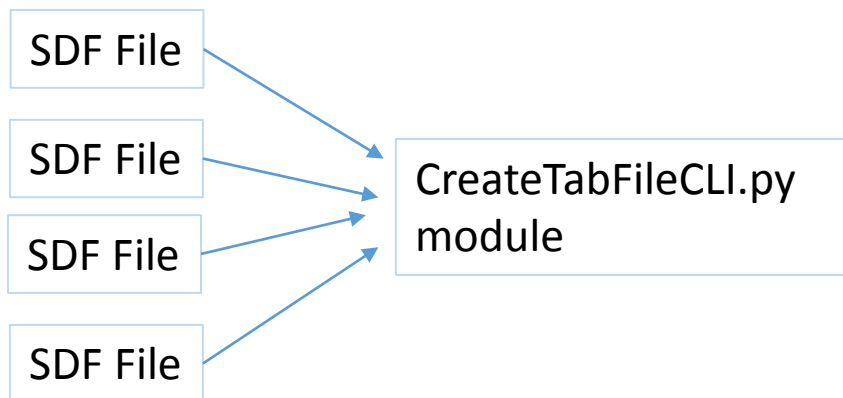
- 2stepReadFromFile.py

>python 2stepReadFromFile.py RankDiff EnergyDiff ColumnSelection

# Περιγραφή Επεξεργασία ½

- Η είσοδος αποτελείται από σειρά αρχείων σε sdf γραμμογραφήση
- Τα αρχεία διαβάζονται (δίνονται ως ορίσματα) από το `createTabFileCLI.py`
- Δημιουργούνται δύο αρχεία εξόδου, ίδιας μορφής. Το ένα αφορά στην στήλη της ενέργειας και το δεύτερο στην στήλη του Rank
- Τα αρχεία αυτά θα αποτελέσουν την είσοδο για τα επόμενα βήματα πραγματοποιείται το φιλτράρισμα και η τελική επιλογή των καλύτερων inhibitors

# Γραμμογράφηση αρχείων εξόδου του createTabFileCLI.py



```
lines 15892
Files 6
BTB 02399 ['n1(c(nc2c(c1=O)cccc2)S)NC(=O)c1cnccc1'] -46.335 -41.646 -45.405 -46.722 -33.548 -37.251
RF 04154 ['n1c(c2c(C(F)(F)F)cc[nH+]c2)nc1c1cccc1'] -36.789 -33.247 -37.949 -42.224 -32.373 -31.239
CD 09023 ['C(=O)(N([C@H](C(=O)OCC)C)Cc1cccc1)Nc1ccc(Cl)cc1'] -47.259 -42.107 -49.051 -44.009 -35.005 -31.175
BTB 12453 ['S(=O)(=O)(c1cc2c(c3c(C2)cc(S(=O)(=O)OCCC(C)C)cc3)cc1)OCCC(C)C'] -49.508 -46.787 -43.647 -42.72 -35.685 -40.526
HTS 07099 ['c1(C(=O)N2CCC(C(=O)[O-])CC2)cc2c(nnn2)cc1'] -39.729 -36.488 -34.929 -46.895 -34.167 -29.477
HTS 13253 ['C1(=Cc2c(nnn(c2)c2cccc2)c2sccc2)C(=O)N(C(=O)N(C1=O)C)C'] -20.297 -48.687 -37.118 -42.437 -36.088 -39.555
HTS 13251 ['C1(=Cc2c(CCl)cccc2)(NH+)1CC[C@H](Cc2cccc2)CC1'] -38.584 -33.897 -38.346 -37.898 -28.739 -30.319
S 07583 ['C([C@H](NC(=O)Cc1c2c(ccc1)cccc2)C(=O)[O-])C(=O)[O-]] -41.247 -36.263 -36.914 -36.729 -30.303 -23.417
NRB 02689 ['c1(cc(=S)c2c(o1)cccc2)c1cccc1'] -37.346 -33.051 -33.753 -39.378 -27.281 -32.079
RJC 00027 ['C\\1(=N\\C(=N)N)/NC(=O)CS1'] -30.732 -30.565 -29.763 -30.27 -27.183 -28.371
RJC 00021 ['n1(c(=S)[nH]nc1C)/N=C\\c1cc(c(c(c1)OC)OC)OC'] -32.617 -36.761 -42.191 -37.595 -29.613 -31.487
KM 01806 ['C(=C\\Nc1ncc(c1)C)/(C(=O)c1sccc1)\\C#N'] -41.172 -37.281 -39.767 -43.801 -36.96 -33.272
RJC 02873 ['c1(n2c(nc1c1cccc1)cccc2)Sc1c(C(=O)OC)cccc1'] -39.419 -38.39 -37.027 -47.281 -23.865 -35.855
KM 01804 ['C(=C\\Nc1ncc(c1)C)/(C(=O)c1ccc(cc1)Cl)\\C#N'] -43.766 -36.131 -38.748 -48.979 -31.733 -34.659
SEW 05144 ['c1(NC(=O)C[C@]23C[C@H]4C[C@H](C2)C[C@H](C3)C4)c(F)cccc1F'] -30.508 -29.117 -29.8 -19.003 -20.774 -27.301
SEW 05146 ['C(=S)(Nc1c(cc(C(F)(F)F)cc1)Cl)N1CCSCC1'] -36.845 -39.09 -36.202 -45.814 -30.181 -32.79
RJC 02878 ['C1(=O)[C@H]2C[C@H](C(=O)[O-])C[C@H]1CC=C2'] -21.356 -19.19 -21.679 -19.535 -13.861 -17.407
RJC 01866 ['c1(nnn1)CC(=O)NNS(=O)(=O)C)c1c(C(F)(F)F)cccc1'] -50.759 -47.128 -47.485 -46.077 -37.563 -36.713
CD 03335 ['n1c(ccc1c1cc([N+](=O)[O-])ccc1)NNC(=O)c1c(Cl)cccc1'] -48.004 -46.742 -50.84 -50.535 -42.807 -43.569
RJC 01992 ['C(=O)(N[C@H](Cl)C(=O)N[C@H](C(=O)[O-])Cc1cccc1)COC(C)C'] -38.952 -36.196 -37.154 -41.198 -32.171 -31.649
XBX 00236 ['n1(c(=O)ccc(c1)Cl)CC(=O)OCC'] -32.112 -31.018 -29.392 -35.2 -28.891 -26.645

lines 15892
Files 6
BTB 02399 ['n1(c(nc2c(c1=O)cccc2)S)NC(=O)c1cnccc1'] -7.138 -5.546 -6.621 -8.56 -5.279 -5.602
RF 04154 ['n1c(c2c(C(F)(F)F)cc[nH+]c2)nc1c1cccc1'] -7.64 -6.122 -7.435 -8.413 -5.41 -5.599
CD 09023 ['C(=O)(N([C@H](C(=O)OCC)C)Cc1cccc1)Nc1ccc(Cl)cc1'] -5.948 -5.957 -7.622 -6.848 -4.169 -4.306
BTB 12453 ['S(=O)(=O)(c1cc2c(c3c(C2)cc(S(=O)(=O)OCCC(C)C)cc3)cc1)OCCC(C)C'] -6.464 -6.307 -6.036 -6.5 -3.607 -4.534
HTS 07099 ['c1(C(=O)N2CCC(C(=O)[O-])CC2)cc2c(nnn2)cc1'] -6.366 -5.265 -4.99 -6.921 -6.142 -5.354
HTS 13253 ['C1(=Cc2c(nnn(c2)c2cccc2)c2sccc2)C(=O)N(C(=O)N(C1=O)C)C'] -4.347 -6.842 -5.21 -5.822 -5.627 -5.229
HTS 13251 ['C1(=Cc2c(CCl)cccc2)(NH+)1CC[C@H](Cc2cccc2)CC1'] -6.793 -4.897 -7.653 -7.63 -4.99 -5.492
S 07583 ['C([C@H](NC(=O)Cc1c2c(ccc1)cccc2)C(=O)[O-])C(=O)[O-]] -6.139 -5.377 -6.989 -6.513 -4.605 -4.113
NRB 02689 ['c1(cc(=S)c2c(o1)cccc2)c1cccc1'] -6.858 -6.854 -6.771 -7.981 -4.939 -6.203
RJC 00027 ['C\\1(=N\\C(=N)N)/NC(=O)CS1'] -5.698 -4.664 -5.622 -6.595 -5.287 -5.528
RJC 00021 ['n1(c(=S)[nH]nc1C)/N=C\\c1cc(c(c(c1)OC)OC)OC'] -6.352 -6.222 -5.903 -6.654 -4.22 -4.707
KM 01806 ['C(=C\\Nc1ncc(c1)C)/(C(=O)c1sccc1)\\C#N'] -6.76 -5.962 -5.699 -7.564 -5.502 -4.847
RJC 02873 ['c1(n2c(nc1c1cccc1)cccc2)Sc1c(C(=O)OC)cccc1'] -5.757 -5.638 -6.047 -8.479 -3.798 -5.576
KM 01804 ['C(=C\\Nc1ncc(c1)C)/(C(=O)c1ccc(cc1)Cl)\\C#N'] -7.187 -6.143 -5.976 -7.559 -5.646 -4.931
SEW 05144 ['c1(NC(=O)C[C@]23C[C@H]4C[C@H](C2)C[C@H](C3)C4)c(F)cccc1F'] -6.031 -5.956 -6.67 -6.208 -3.778 -4.707
SEW 05146 ['C(=S)(Nc1c(cc(C(F)(F)F)cc1)Cl)N1CCSCC1'] -6.816 -6.774 -5.921 -8.641 -5.004 -5.818
RJC 02878 ['C1(=O)[C@H]2C[C@H](C(=O)[O-])C[C@H]1CC=C2'] -5.3 -5.912 -5.845 -5.51 -3.892 -5.485
RJC 01866 ['c1(nnn1)CC(=O)NNS(=O)(=O)C)c1c(C(F)(F)F)cccc1'] -7.698 -7.284 -7.333 -7.048 -5.193 -5.566
CD 03335 ['n1c(ccc1c1cc([N+](=O)[O-])ccc1)NNC(=O)c1c(Cl)cccc1'] -6.517 -6.817 -7.304 -6.621 -5.73 -5.936
```

```
lines 15892
Files 6
BTB 02399 ['n1(c(nc2c(c1=O)cccc2)S)NC(=O)c1cnccc1'] -7.138 -5.546 -6.621 -8.56 -5.279 -5.602
```

# Φιτράρισμα σύμφωνα με Rank και Energy Diff

Στο δεύτερο επίπεδο χρησιμοποιείται το 2stepReadFromFile.py στο οποίο η είσοδος είναι η έξοδος του προηγούμενου βήματος.

Ο χρήστης εισάγει παραμετρικά τα εξής

Rank diff

Energy diff

Base File

Το εργαλείο επιστρέφει δύο αρχεία (energy και rank) με τα compounds που ικανοποιούν τις συνθήκες

# Γραμμογράφηση αρχείων εξόδου του 2stepReadFromFile.py

[illegible]

## 2stepReadFromFile.py module

NRB 01104	['[N+](=O)c1cc(c(/N=N/c2c(cc(c2)[NH+](CCCCCCC)CC)NC(=O)C)c(C#N)c1(C#N)[O-]']	-38.6647922992706	-55.3457403182983	-44.851330280304	-42.9705014228821	-38.6371648311615	-40.5034439563751
NRB 03993	['[C@]12(N([C@H](C(S1)(C)C)C(=O)OC2)C(=O)C)[C@H]([C@H](OC(=O)C)COC(=O)C)OC(=O)C']	-29.2555182278156	-43.027473449707	-26.0412139892578	-27.9629883766174	0	-26.9252021312714
HTS 02534	['n1(c(nc2c(c1=O)cc(c2)OC)OC)SCC(=O)Nc1cc(c(cc1)F)Cl1c1cccc1']	-45.3454891443253	-57.2409391403198	-46.9882535934448	-45.8716766834259	-37.3092696666718	-45.2379838228226
EN 00092	['c1(c(C(=O)Oc2cc(nn2C)C(F)(F)F)c(on1)C)c1c(Cl)cccc1Cl']	0	-40.2610639333725	-16.3826662302017	0	-21.3230545520782	-27.7172078490257
SEW 03500	['c1(S(=O)(=O)Oc2cc(ccc2)C)c(sc(c1c1cccc1)C(F)(F)F)C(=O)OC']	0	-40.3502403497696	-24.4344091415405	0	-22.7187485694885	-15.1760445833206
SEW 02121	['C(=C\\c1cccc1)(\\C(=O)Oc1ccc(F)cc1)/c1ccsc1']	-18.3019876480103	-41.7216922938824	-26.5644154548645	0	0	0
NRB 03698	['[C@]12(C(=CC[C@H]1[C@H]1[C@H]([C@]3([C@H](CC1)C[C@H](OC(=O)CC)CC3)C)C(=O)C2)C(=O)C)C']	-29.051701426506	-41.1210403442383	-30.9793671369553	0	-22.2917002439499	-26.2635596990585



# Μέρος Δεύτερο Web Interface

- Για τις ανάγκες της εφαρμογής αναπτύχθηκε ένα πλήρως λειτουργικό Web Interface
- Το περιβάλλον είναι Ubuntu linux, Apache Web Server, HTML5, Python

# Οθόνες Εφαρμογής Login, Main Form

ReRank VS results

**Login**

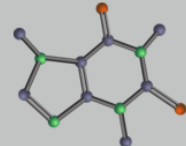
UserName

Password

ReRank VS results

- 
- 
- 
- 

Select from the menu on the left panel

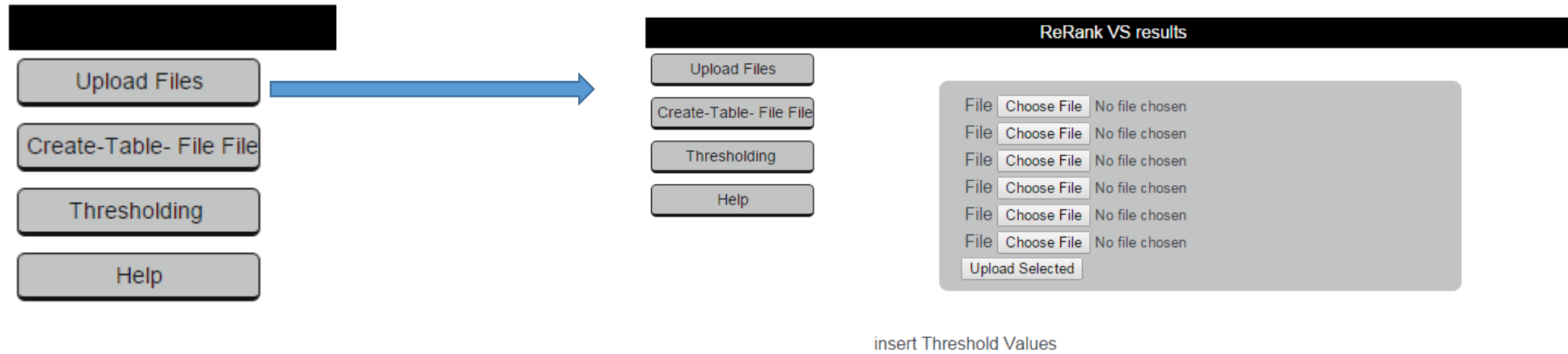


- 
- 
- 
- 

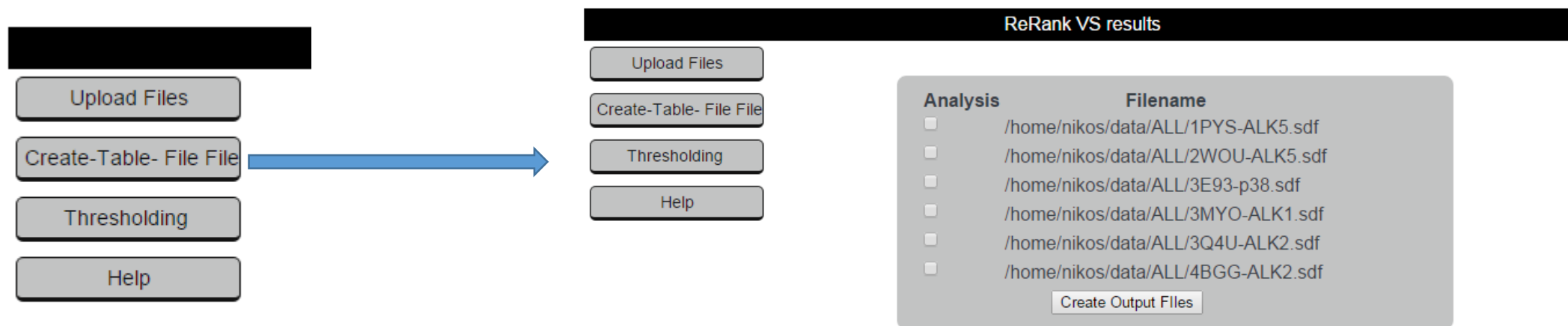
Nikos Perdikopanis-Biomolecules Modeling-Final Project 2016

Nikos Perdikopanis-Biomolecules Modeling-Final Project 2016

# Upload Files

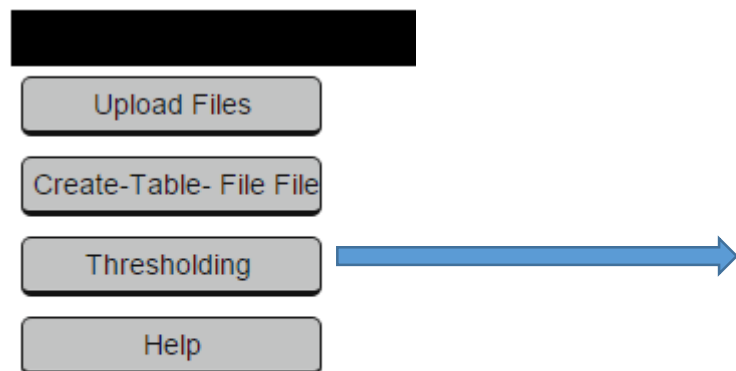


# Create Table-File



insert Threshold Values

# Thresholding 1/2



ReRank VS results

Upload Files

Create-Table- File File

Thresholding

Help

Rank Threshold

Energy Threshold

Filename	Base File
/home/nikos/data/ALL/1PYS-ALK5.sdf	<input type="radio"/>
/home/nikos/data/ALL/2WOU-ALK5.sdf	<input type="radio"/>
/home/nikos/data/ALL/3E93-p38.sdf	<input type="radio"/>
/home/nikos/data/ALL/3MYO-ALK1.sdf	<input type="radio"/>
/home/nikos/data/ALL/3Q4U-ALK2.sdf	<input type="radio"/>
/home/nikos/data/ALL/4BGG-ALK2.sdf	<input type="radio"/>

Go on

# Thresholding 2/2

Rank Threshold

Energy Threshold

**Filename** **Base File**

/home/nikos/data/ALL/1PYS-ALK5.sdf ☒

/home/nikos/data/ALL/2WOU-ALK5.sdf ☐

/home/nikos/data/ALL/3E93-p38.sdf ☐

/home/nikos/data/ALL/3MYO-ALK1.sdf ☐

/home/nikos/data/ALL/3Q4U-ALK2.sdf ☐

/home/nikos/data/ALL/4BGG-ALK2.sdf ☐



lines 15892 Files 6

ReRank VS results

Filtering for  
Rank Threshold=2  
Energy Threshold=2  
Column Data=1

RJC 00945	[*C@@]12(c3ccc([N+](=O)[O-])cc3)C[C@H]3C[C@@H](C1)C[C@@H](C2)C3]	-7.002	-3.146	-4.716	-4.415	-4.763	-4.02
BTB 04693	[*S(=O)(=O)N1CC([NH+](c2c([N+](=O)[O-])cc(C(F)(F)F)cc2)CC1)c1cc(c(cc1)OC)OC]	-7.001	-3.68	-4.645	-4.672	-4.261	-4.304
RJC 02314	[*c12c(c3nc4c(cc3)cccc4)c(n(c(=O)c1cc(cc2)[N+](=O)[O-])Cc1cccc1)N]	-9.278	-6.347	-6.64	-5.455	-4.99	-6.444
JFD 00854	[*c1(c(=O)n(c2c(c1C(C(=O)OCC)C(=O)OCC)cccc2)c1cccc1)[N+](=O)[O-]]	-7.339	-4.826	-4.554	100.0	-3.06	-4.241
SEW 01857	[*c1(C(=O)NC(=[NH2+])[C@]23C[C@@H]4C[C@H](C2)C[C@H](C3)C4)c(oc1C)C]	-6.874	-4.717	-3.835	-4.732	-3.252	-4.309
CD 08680	[*S(=O)(=O)/N=C/NNC(=O)c1c(ncc1C)c1c(Cl)cccc1Cl)\SC)c1cccc1]	-5.923	-3.664	-3.375	-3.222	-0.637	-3.446
BTB 06638	[*S(=O)(=O)c1cc(c([N+](=O)[O-])cc1)Oc1ncc([N+](=O)[O-])cc1c1cccc1]	-8.284	-5.763	-5.117	-4.791	-3.165	-5.344
BTB 02647	[*n1(c(=N)sc1c1ccc(cc1)Cl)c1c(ccc1)[NH+]1CCOCC1]	-6.96	-4.377	-3.833	100.0	-3.565	-4.632
HTS 03117	[*c1(c([n+](c[nH]1)C)N1CC([NH+](C(c2cccc2)c2cccc2)CC1)[N+](=O)[O-])	-7.355	-4.729	-4.253	-3.411	-2.654	-3.768
NRB 03809	[*C@]12([C@@H]([C@@H]3[C@@H]([C@](C1)(O)C)c1c(cc(OC(=O)C)cc1)CC3)CC[C@H]2OC(=O)C)C]	-6.417	-4.4	-4.295	-4.395	-4.24	-3.379
HTS 09909	[*n12c(c(c3c(c1Sc1[nH]c4c(cccc4)[nH+]1)CCCC3)C#N)[nH+][c1c2cccc1]	-6.684	-5.638	-5.422	-5.106	-3.686	-5.167
BTBG 00203	[*C@@]12(c3ccc(cc3)OC)C[C@H]3C[C@@H](C1)C[C@@H](C2)C3]	-6.65	-4.241	-4.557	-4.346	-4.364	-3.894
S 03966	[*N+](=O)NC(=[NH2+])NC(C#C)C[C@H](O-)]	-5.212	-2.924	-3.172	-2.779	-2.924	-2.31
CD 08935	[*C@@]12(C(=O)N3CCSCC3)C[C@@H]3C[C@H](C2)C[C@H](C1)C3]	-6.65	-4.469	-3.879	-4.402	100.0	-4.012
S 13684	[*C12=C([C@@H](C(=C(O1)N)C#N)c1ccc(SC)c1)CCc1c2cccc1]	-7.766	-5.052	-5.484	-5.187	-3.762	-5.293
HTS 07836	[*N(c1c(ccc1)[NH+]1CCOCC1)C(=O)C[N@@H+](C1)CC[C@@H](C(=O)N)CC1]	-9.207	-6.97	-6.598	-5.277	-3.918	-6.015

# Για υλοποίηση

- Ολοκλήρωση module μεταφόρτωσης αρχείων
- Επιλογή υποσυνόλου αρχείων για ανάλυση
- Ενοποίηση φιλτραρίσματος (Rank και energy)