VI-SEEM Service Access Application Form - 1st Call

Survey response 1

Response ID
74

Date submitted
2016-11-14 15:29:40

Last page
17

Start language
en

Date started
2016-11-14 14:12:47

Date last action
2016-11-14 15:29:40

IP address
195.251.115.2

Referrer URL

Total time
4616.91

General Project Information

Project name
Predicting Farnesoid X receptor (FXR) - inhibitor structures and affinities for computer-aided drug design within the D3R challenge

Project Acronym
D3R

Project Description
Computational tools, such as docking and scoring, for modeling the interactions of proteins with drug-like compounds (ligands) hold great promise to speed the discovery of new, safer medications and reduce the cost of the drug discovery process. However, there is still a need for improved methods of predicting ligand poses and affinities or relative affinities. In this project, we were provided by a blinded unpublished dataset ,within the D3R challenge, containing high quality crystal structures and potency data for testing and improving ligand-protein docking algorithms and their scoring protocols. It is based on the Farnesoid X receptor (FXR) target; the dataset is kindly contributed by Roche and curated by D3R. The FXR nuclear receptor forms a heterodimer with RXR when activated, and binds to hormone response elements on DNA, leading to up- or down-regulation of the expression of certain genes. FXR agonists are regarded as potential therapeutics for dyslipidemia and diabetes. The blind dataset has 36 crystal structures with resolution < 2.6Å; and binding data (IC50s) for 102 compounds across five orders of magnitude (Scintillation Proximity Assay), comprising four chemical series and 6 miscellaneous compounds. We have been provided only with the structures of the 102 ligands and using publicly available data about FXR we are called to : a) Predict the crystallographic poses of 36 ligands spanning all chemical series, b) Predict affinities, or affinity rankings, for these ligands, and also for the other 66 ligands, c) Predict the relative binding affinities for two designated free energy subsets of 18 and 15 compounds. This project aims to help advance such computational technologies as docking and scoring and free energy methods by collecting and enhancing ligand-protein datasets, and using them as a basis for community-wide, blinded prediction challenges.

Research Area
LS Area B: Computer-aided drug design.

Group time: General Project Information
1457.49

Principal Investigator

Title
Dr

First Name
Zoe

Last Name
Cournia

Email
zcournia@bioacademy.gr

Gender (Male/Female)
Female

Date of Birth
1977-11-07 00:00:00

Nationality
Hellenic

Phone number
+302106597195

Organisation and job title

Position held
Researcher C

Organization Name
Biomedical Research Foundation Academy of Athens

Department
Translational Research

Research Group Name
Molecular Modeling and Drug Design

Street Address
Soranou Ephessiou 4

City
Athens

Postal Code
11527

Country (of PI institute)
Greece

Do you want to add a Collaborator
Yes

Group time: Principal Investigator
120.81

Project Collaborator 1

Title
Dr

First Name
Sofia

Last Name
Vasilakaki

Email
svasilak@chem.uoa.gr

Gender (Male/Female)
Female

Date of Birth
1985-03-16 00:00:00

Nationality
Hellenic

Phone number
+306988726101

Organization and job title

Position held
Postdoctoral Associate

Organization Name
BRFAA

Department
Translational Research

Street Address
Soranou Ephesiou 4

Postal Code
11527

Country
Greece

Do you want to add another Collaborator?
Yes

Group time: Project Collaborator 1
133.12

Project Collaborator 2

Title
Ms

First Name
Christina

Last Name
Athanasiou

Email
chrisath93@hotmail.gr

Gender (Male/Female)
Female

Date of Birth
1994-01-19 00:00:00

Nationality
Hellenic

Phone number
+306976906918

Organizationa and job title

Position held
Diploma thesis student

Organization Name
BRFAA

Group
Molecular Modeling and Drug Design

Department
Translational Research

Street Address
Soranou Ephesiou 4

City
Athens

Postal Code
11527

Country
Greece

Do you want to add another Collaborator?
Yes

Group time: Project Collaborator 2
130.42

Project Collaborator 3

Title
Dr

First Name
Thomas

Last Name
Steinbrecher

Email
thomas.steinbrecher@schrodinger.com

Gender (Male/Female)
Male

Date of Birth
1974-04-21 00:00:00

Nationality
German

Phone number
+49 621 438 55 170

Organization and job title

Position held
Application Scientist

Organization Name
Schrodinger

Group
Mannheim

Department
Research

Street Address
Dynamostrasse 13

City
Mannheim

Postal Code
D-68165

Country
Germany

Do you want to add another Collaborator?
No

Group time: Project Collaborator 3
110.98

Project Collaborator 4

Title

First Name

Last Name

Email

Date of Birth

Nationality

Phone Number

Organization and job title

Position held

Organization Name

Group

Department

Street Address

City

Postal Code

Country

Do you want to add another Collaborator?
N/A

Project Collaborator 5

Title

First Name

Last Name

Email

Date of Birth

Nationality

Phone number

Organization and job title

Position held

Organization Name

Group

Street Address

City

Postal Code

Country

Abstract of the project

If the project is successful this will be published on VI-SEEM website unless you mark it as confidential below. Please make this summary understandable to a general audience. (Maximum 500 words)
Computational tools, such as docking and scoring, for modeling the interactions of proteins with drug-like compounds (ligands) hold great promise to speed the discovery of new, safer medications and reduce the cost of the drug discovery process. However, there is still a need for improved methods of predicting ligand poses and affinities or relative affinities. In this project, we were provided by a blinded unpublished dataset,within the D3R challenge, containing high quality crystal structures and potency data for testing and improving ligand-protein docking algorithms and their scoring protocols. It is based on the Farnesoid X receptor (FXR) target; the dataset is kindly contributed by Roche and curated by D3R. The FXR nuclear receptor forms a heterodimer with RXR when activated, and binds to hormone response elements on DNA, leading to up- or down-regulation of the expression of certain genes. FXR agonists are regarded as potential therapeutics for dyslipidemia and diabetes. The blind dataset has 36 crystal structures with resolution < 2.6Å; and binding data (IC50s) for 102 compounds across five orders of magnitude (Scintillation Proximity Assay), comprising four chemical series and 6 miscellaneous compounds. We have been provided only with the structures of the 102 ligands and using publicly available data about FXR we are called to : a) Predict the crystallographic poses of 36 ligands spanning all chemical series, b) Predict affinities, or affinity rankings, for these ligands, and also for the other 66 ligands, c) Predict the relative binding affinities for two designated free energy subsets of 18 and 15 compounds. This project aims to help advance computational technologies such as docking and scoring and free energy methods by using unknown ligand-protein datasets, predicting their outcome and helping advance the state of the art of the computer-aided drug design techniques.

Is the summary above confidential?
No

Group time: Abstract of the project
124.67

Required VI-SEEM Services

Specify the generic VI-SEEM services that your application requires [HPC Resources]
Yes

Specify the generic VI-SEEM services that your application requires [Grid Resources]
No

Specify the generic VI-SEEM services that your application requires [Cloud Resources]
No

Specify the generic VI-SEEM services that your application requires [VI-SEEM simple storage service]
No

Specify the generic VI-SEEM services that your application requires [VI-SEEM repository service]
No

Specify the generic VI-SEEM services that your application requires [VI-SEEM archival service]
No

Specify the generic VI-SEEM services that your application requires [None of these]
No

Specify which, if any, application specific services your application can/will use: [Climate Application Specific Service – Live Access Sever (LAS) – http://las.vi-seem.eu]
No

Specify which, if any, application specific services your application can/will use: [CH Application Specific Service - Clouder - http://dchrepo.vi-seem.eu]
No

Specify which, if any, application specific services your application can/will use: [Life Application Specific Service – ChemBioServer - http://bioserver-3.bioacademy.gr/Bioserver/ChemBioServer/]
Yes

Specify which, if any, application specific services your application can/will use: [None of these]
No

Group time: Required VI-SEEM Services
19.02

HPC Resources - Codes and computational resources requested

HPC CODE 1 Details

CODE 1 - NAME
Schrodinger FEP+/Desmond GPU

CODE 1 - URL (if it exists)
https://www.schrodinger.com/

CODE 1 - Licence (Type and URL)
https://www.schrodinger.com/

Did/Do you develop CODE 1?
No

Do you want to add another code?
No

HPC CODE 2 Details

CODE 2 - NAME

CODE 2 - URL (if it exists)

CODE 2 - Licence (Type and URL)

Did/Do you develop CODE 2?
N/A

Do you want to add another code?
N/A

HPC CODE 3 Details

CODE 3 - NAME

CODE 3 - URL (if it exists)

CODE 3 Licence (Type and URL)

Did/Do you develop CODE 3?
N/A

Overall HPC Compute Requirements for all the codes that you have declared

Number of total core-hours required to run all the codes on CPU only nodes (in hours)
0

Number of total core-hours required to run all the codes in GPU nodes (in hours)
1500

Number of total core-hours required to run all the codes in Phi nodes (in hours)
0

Wall-clock time of a typical job execution (in hours)
72

Are you able to write checkpoint during the job execution?
No

Expected single job size (number of cores) and single job memory (total memory usage over all cores of a single job)

Minimum (number of cores) for a single job
10

Minimum (total memory usage across all nodes used) for a single job (in GB)
20

Maximum (number of cores) for a single job (in GB)
22

Maximum (total memory usage across all nodes used) for a single job
20

Storage
60 GB

Total storage (work) - result and large input files (GB/TB)
500 GB

Total storage (home) / source code and scripts (GB/TB)
1 GB

Data transfer

Total amount of data to be transferred to/from the production system (GB)
500 GB

Group time: HPC Resources - Codes and computational resources requested
210.74

Grid Resources - Codes and computational resources requested

Grid CODE 1 Details

CODE 1 - NAME

CODE 1 - URL (if it exists)

CODE 1 - Licence (Type and URL)

Did/Do you develop CODE 1?
N/A

Do you want to add another code?
N/A

Grid CODE 2 Details

CODE 2 - NAME

CODE 2 - URL (if it exists)

CODE 2 - Licence (Type and URL)

Did/Do you develop CODE 2?
N/A

Do you want to add another code?
N/A

Grid CODE 3 Details

CODE 3 - NAME

CODE 3 - URL (if it exists)

CODE 3 Licence (Type and URL)

Did/Do you develop CODE 3?
N/A

Overall Grid Compute Requirements for all the codes that you have declared

Number of total jobs that you are planning to submit using all codes you have specified above

Wall-clock time of a typical job execution (in hours)

Minimum memory usage for a single job (in MB)

Maximum memory usage for a single job (in MB)

Storage

Total storage (work) - result and large input files (GB/TB)

Data transfer

Total amount of data to be transferred to/from the production system (GB)

Cloud Resources requested

Total number of required virtual cores required

Total number of VMs required

Total amount of file storage space required (in GB), i.e. Storage space provided in attached volumes

Total amount of Hard Disk Storage required (in GB). i.e. The total ammount of the boot disks of all VMs that you require

Total amount of RAM required(in GB) i.e. overall ammount of RAM over all VMs that are required

Total number of public IPv4 addresses requred

Operating System(s) required [Linux Distribution]
N/A

Operating System(s) required [A version of the Microsoft Windows OS]
N/A

VI-SEE Repository Service requirements

Total amount of sotrage space needed for your data (in GB)

VI-SEE Archival Service requirements

Total amount of sotrage space needed for your data (in GB)

Application Specific LAS Service

Total amount of storage space needed for your shared data (in GB). If you will not use LAS to share data please fill in the box with 0.

Application Specific Clowder Service

Total amount of storage space needed for your data storage (in GB). If you will not use LAS to store data please fill in the box with 0.

Please provide us with information on the data format of the dataset to be provided to Clowder

Scientific Case

Describe your research project. Include discussion of the scientific questions that you are planning to address and the overall scientific goals of the project. It is important that you describe the novelty, impact and timeliness of the proposal.

Introduction

Computational tools, such as docking and scoring, for modeling the interactions of proteins with drug-like compounds (ligands) hold great promise to speed the discovery of new, safer medications and reduce the cost of the drug discovery process. However, there is still a need for improved methods of predicting ligand poses and affinities or relative affinities.

In this project, we were provided with 102 inhibitors of the Farnesoid X receptor (FXR) target, courtesy of Roche Pharmaceuticals. 36 of these ligands have co-crystal structures but these have not yet been publicly disclosed. The FXR nuclear receptor forms a heterodimer with RXR when activated, and binds to hormone response elements on DNA, leading to up- or down-regulation of the expression of certain genes. FXR agonists are regarded as potential therapeutics for dyslipidemia and diabetes.

This blind dataset has 36 crystal structures with resolution < 2.6Å; and binding data (IC50s) for 102 compounds across five orders of magnitude (Scintillation Proximity Assay), comprising four chemical series and 6 miscellaneous compounds. By using publicly available data about FXR we are called to : a) Predict the crystallographic poses of 36 ligands spanning all chemical series, b) Predict affinities, or affinity rankings, for these ligands, and also for the other 66 ligands, c) Predict the relative binding affinities for two designated free energy subsets of 18 and 15 compounds.Our task is to predict the ligand-proriten complexes using pink

This project aims to help advance computational technologies such as docking and scoring and free energy methods by using unknown ligand-protein datasets, predicting their outcome and helping further the state of the art of the computer-aided drug design techniques by improving ligand-protein docking algorithms and their scoring protocols.

The D3R Challenge

The challenge involves a protein-ligand dataset for the farnesoid x receptor (FXR) target, generously donated by Roche Pharmaceuticals. The dataset comprises:
o IC50 data for 102 compounds in total, 96 in four chemical series (benzimidazoles, isoxazoles, spiros and sulfonamides) and 6 miscellaneous compounds.
o Potency range of 0.000335 – 62.37 μM for 92 compounds, and 10 having potency > 100 μM.
o 36 co-crystal structures and one apo, with representatives from each of the four chemical series. Resolutions range from 1.8 – 2.6 Å.

This challenge involves two stages. Stage 1 of the Challenge is to predict the ligand poses of the available crystal structures and also to predict or rank the potencies of all ligands, including those for which crystal structures are not available. After Stage 1 has closed, all available co-crystal structures will be made public. The Stage 2 Challenge is to repeat the affinity predictions or rankings, this time using the additional disclosed ligand-pose information.

Compounds whose IDs are listed in Appendix A were prepared and tested as 50:50 racemic mixtures. However, available co-crystal structures and discussions with Roche scientists indicate that the S isomer is much more active than the R. The provided SMILES are the S form but you're free to consider the R. You'll be evaluated against the uncorrected experimental IC50s.

Two subsets have been selected, 15 sulfonamides and 18 spiros, which contain chemically similar compounds and thus lend themselves to the calculation of relative binding affinities by alchemical methods, such as free energy perturbation. The compound IDs for these sets are provided in Appendix B and C, at the end of these instructions. The free energy (FE) challenge can be valid for not only Stage 1 but also Stage 2.

Factors that make this Challenge interesting:

o There are two helices adjacent to the ligand binding site that can adopt varied conformations. The conformations observed in the blinded dataset are well exemplified in publicly available co-crystal structures of FXR in the Protein Data Bank.

o There are PDB entries in the Protein Data Bank with ligands in the benzimidazole1,2 and isoxazole3 chemical series, but co-crystal structures with the spiros or sulfonamides are not publicly available, as of today.

o Water mediated protein interactions are important for the binding of some but not all ligands.

o Some ligands have rings with nontrivial puckering options.

Binding assay and crystallization conditions
The FXR binding affinities were carried out using the Scintillation Proximity Assay (SPA)4, a radioligand displacement assay. The assay buffer contained 50 mM HEPES (pH 7.4), 10 mM NaCl, 5 mM $MgCl_2$ and 0.01% CHAPS. The reactions were incubated for

30 min in the presence of [3H]2,N-dicyclohexyl-2-[2-(2,4-dimethoxy-phenyl)-benzoimidazol-1-yl]acetamide, the test compound and the buffer. The amount of radioligand that remained bound was determined; dose response curves were then generated and the IC50s calculated.

Different crystallization conditions had to be used for the diverse chemotypes; the crystallization conditions for the compound IDs for the pose prediction challenge are provided as a csv file named Data_set_fxr_crystallization_conditions.csv.

Note that, per Roche, the crystallization solutions for FXR_10 and FXR_26 were unbuffered, and their pHs are not known.

Predicted Aggregators
All compounds have been subjected to Open Eye FILTER (http://www.eyesopen.com/filter) to identify known and predicted aggregators. None of the Challenge compounds are known aggregators, but 82 out of the 102 compounds are “predicted” as aggregators using the QSAR model within FILTER. Whether they are true aggregators or not has not been tested.

Due dates
Our predictions (poses, ligand affinities/rankings and/or FE set predictions) for Stage 1 must be uploaded to the D3R website by 5pm (PST), 21st November 2016. The experimental ligand-protein poses will be released immediately after Stage 1 closes.

The Stage 2 predictions (ligand affinities/ rankings) are due by 5pm (PST), 1st February 2017. The experimental ligand-protein IC50s will be released immediately after Stage 2 closes.

Strategy for computer-aided prediction of protein-ligand structures and inhibitor affinities

1) Methodology for predicting the poses for the 36 ligands

Following library and receptor preparation for the 102 provided ligands, each compound in this library will be virtually docked into the target binding site with the program GLIDE. Docking aims to predict the ligand-protein complex structure by exploring the conformational space of the ligands within the binding site of the protein. A scoring function is then utilized to approximate the free energy of binding between the protein and the ligand in each docking pose. Docking and scoring will produce the final structures of the complexes. The crystal structure in which each ligand will be docked, will be carefully selected by comparing the similar of the native ligands structure and its interaction fingerprints with the binding pocket, compared to the unknown ligand. Once the crystal structure for docking will be selected, the final, predicted pose of the ligand will be chosen based on similarity with the native legend (flexible docking alignment) and Metadynamics calculations to choose between two or more similar poses.

2) Methodology for predicting the affinities, or affinity rankings, for all 102 compounds.

Based on available experimental results and docking scores, the ranking of the different classes of inhibitors will first be predicted (spires, benzimidazoles, sulfonamides, miscellaneous). Then, Free energy perturbation (FEP) calculations, which can yield very accurate relative free energies of binding, will be performed to identify the ranking within each subgroup of inhibitors (see also below).

3) Methodology for predicting therelative binding affinities (in kcal/mol) for the free energy subsets of 18 and 15 compounds.

Computer-aided optimization of small molecule drugs has revealed major successes due to improvements in the accuracy of predicted protein ligand binding affinities. Free energy perturbation (FEP) calculations can yield very accurate relative free energies of binding (~1kcal/mol error). We will perform FEP calculations coupled with Molecular Dynamics (MD) simulations for the FXR ligands as described above. This system consists of ~50000 atoms. The simulation of such large biomolecular systems and the need to accelerate the production of the FEP calculation is computationally demanding. Thus, a significant computation power, as the one that is provided by HPC resources, is needed. The usage of GPUs is of high importance, since 3 perturbations per day that run on 4 GPUs, would require 768 cores if it was to run on CPUs, thus demanding valuable resources.

Evaluation of predictions
Pose predictions will be evaluated based on, at minimum, symmetry-corrected RMSD to crystallographic conformations. Additional criteria may be based on ligand-protein contacts and/or overlap of predicted and experimental electron densities. We also might evaluate predicted conformational changes of the protein binding site. Affinity predictions/rankings will be evaluated based on, at minimum, accuracy of ranking.

For more information, please see:
https://drugdesigndata.org/about/grand-challenge-2

Appendix A. Compounds experimentally tested as 50:50 racemic mixtures
FXR_37, FXR_39, FXR_40, FXR_42, FXR_50, FXR_51, FXR_52, FXR_53, FXR_54, FXR_55, FXR_56, FXR_57, FXR_58, FXR_59, FXR_60, FXR_61, FXR_62, FXR_63, FXR_64, FXR_66, FXR_67, FXR_68, FXR_69, FXR_70, FXR_71, FXR_72.
Appendix B. FE Set 1
FXR_17□□CCOC(=O)c1ccc(NC(=O)c2c3CN(CCc3nn2c4ccccc4)S(=O)(=O)c5cccs5)cc1
FXR_45□□CCOC(=O)c1ccc(NC(=O)c2c3CN(CCc3nn2c4ccc(OC(F)(F)F)cc4)S(=O)(=O)c5cccs5)cc1
FXR_46□□NC(=O)c1ccc(NC(=O)c2c3CN(CCc3nn2c4ccccc4)S(=O)(=O)c5cccs5)cc1
FXR_47□□CCOC(=O)c1cccc(NC(=O)c2c3CN(CCc3nn2c4ccccc4)S(=O)(=O)c5cccs5)c1
FXR_48□□CCOC(=O)Cc1ccc(NC(=O)c2c3CN(CCc3nn2c4ccccc4)S(=O)(=O)c5cccs5)cc1
FXR_49□□CC(=O)c1ccc(NC(=O)c2c3CN(CCc3nn2c4ccccc4)S(=O)(=O)c5cccs5)cc1
FXR_91□□O=C(Nc1ccccc1)c2c3CN(CCc3nn2c4ccccc4)S(=O)(=O)c5cccs5
FXR_93□□O=C(Nc1ccccc1)n2c3CN(CCc3cc2c4ccccc4)S(=O)(=O)c5cccs5
FXR_95□□CC(=O)Nc1ccc(NC(=O)c2c3CN(CCc3nn2c4ccccc4)S(=O)(=O)c5cccs5)cc1
FXR_96□□CN(C)C(=O)c1ccc(NC(=O)c2c3CN(CCc3nn2c4ccccc4)S(=O)(=O)c5cccs5)cc1
FXR_98□□CNC(=O)c1ccc(NC(=O)c2c3CN(CCc3nn2c4ccccc4)S(=O)(=O)c5cccs5)cc1
FXR_99□□COc1ccc(NC(=O)c2c3CN(CCc3nn2c4ccccc4)S(=O)(=O)c5cccs5)cc1
FXR_100□NS(=O)(=O)c1ccc(NC(=O)c2c3CN(CCc3nn2c4ccccc4)S(=O)(=O)c5cccs5)cc1
FXR_101□OC(=O)c1ccc(NC(=O)c2c3CN(CCc3nn2c4ccccc4)S(=O)(=O)c5cccs5)cc1
FXR_102□O=C(Nc1ccc(cc1)C(=O)N2CCOCC2)c3c4CN(CCc4nn3c5ccccc5)S(=O)(=O)c6cccs6

Appendix C. FE Set 2
FXR_10□□OC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4cccs4)c5cc(Br)ccc25)cc1
FXR_12□□OC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4ccccc4Cl)c5cc(Br)ccc25)cc1
FXR_38□□COC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4cccs4)c5cc(Br)ccc25)cc1
FXR_41□□COC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4ccccc4Cl)c5cc(Br)ccc25)cc1
FXR_73□□Oc1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4cccs4)c5cc(Br)ccc25)cc1
FXR_74□□OC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4ccccc4Br)c5cc(Br)ccc25)cc1
FXR_75□□Brc1ccc2N(Cc3ccncc3)C(=O)C4(CCN(CC4)S(=O)(=O)c5cccs5)c2c1
FXR_76□□OC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4ccccc4)c5cc(Br)ccc25)cc1
FXR_77□□OC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4cccc(Cl)c4Cl)c5cc(Br)ccc25)cc1
FXR_78□□OC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4c(Cl)cccc4Cl)c5cc(Br)ccc25)cc1
FXR_79□□OC(=O)c1cccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4cccs4)c5cc(Br)ccc25)c1
FXR_81□□Cc1c(Cl)cccc1S(=O)(=O)N2CCC3(CC2)C(=O)N(Cc4ccc(cc4)C(=O)O)c5ccc(Br)cc35
FXR_82□□OC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4cccc(Cl)c4F)c5cc(Br)ccc25)cc1
FXR_83□□OC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4cc(Cl)ccc4Cl)c5cc(Br)ccc25)cc1
FXR_84□□OC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4ccccc4F)c5cc(Br)ccc25)cc1
FXR_85□□Cc1ccccc1S(=O)(=O)N2CCC3(CC2)C(=O)N(Cc4ccc(cc4)C(=O)O)c5ccc(Br)cc35
FXR_88□□OC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4ccccc4C(F)(F)F)c5cc(Br)ccc25)cc1
FXR_89□□OC(=O)c1ccc(CN2C(=O)C3(CCN(CC3)S(=O)(=O)c4ccc(Cl)cc4)c5cc(Br)ccc25)cc1

Recent bibliographic references that are relevant to the project.

(1)□Richter, H. G.; Benson, G. M.; Bleicher, K. H.; Blum, D.; Chaput, E.; Clemann, N.; Feng, S.; Gardes, C.; Grether, U.; Hartman, P.; Kuhn, B.; Martin, R. E.; Plancher, J. M.; Rudolph, M. G.; Schuler, F.; Taylor, S. Bioorganic & medicinal chemistry letters 2011, 21, 1134.
□(2)□Richter, H. G.; Benson, G. M.; Blum, D.; Chaput, E.; Feng, S.; Gardes, C.; Grether, U.; Hartman, P.; Kuhn, B.; Martin, R. E.; Plancher, J. M.; Rudolph, M. G.; Schuler, F.; Taylor, S.; Bleicher, K. H. Bioorganic & medicinal chemistry letters 2011, 21, 191.
□(3)□Feng, S.; Yang, M.; Zhang, Z.; Wang, Z.; Hong, D.; Richter, H.; Benson, G. M.; Bleicher, K.; Grether, U.; Martin, R. E.; Plancher, J. M.; Kuhn, B.; Rudolph, M. G.; Chen, L. Bioorganic & medicinal chemistry letters 2009, 19, 2595.
□(4)□Gardes, C.; Blum, D.; Bleicher, K.; Chaput, E.; Ebeling, M.; Hartman, P.; Handschin, C.; Richter, H.; Benson, G. M. Journal of lipid research 2011, 52, 1188.
(5) Repasky MP, Shelley M, Friesner RA. Flexible ligand docking with Glide. Curr Protoc Bioinformatics. 2007 Jun;Chapter 8:Unit 8.12.
(6) Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem. 2004 Mar 25;47(7):1750-9.
(7) Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. 2004 Mar 25;47(7):1739-49.
(8) Mansha M, Kumari UU, Cournia Z, Ullah N. Pyrazole-based potent inhibitors of GGT1: Synthesis, biological evaluation, and molecular docking studies. Eur J Med Chem. 2016 Nov 29;124:666-676.
(9) Ioannidis H, Drakopoulos A, Tzitzoglaki C, Homeyer N, Kolarov F, Gkeka P, Freudenberger K, Liolios C, Gauglitz G, Cournia Z, Gohlke H, Kolocouris A. Alchemical Free Energy Calculations and Isothermal Titration Calorimetry Measurements of Aminoadamantanes Bound to the Closed State of Influenza A/M2TM. J Chem Inf Model. 2016 May 23;56(5):862-76.
(10) Lionta E, Spyrou G, Vassilatis DK, Cournia Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. Curr Top Med Chem. 2014;14(16):1923-38.

Describe the numerical methods and algorithms that you are planning to use, improve, or develop. List the codes, packages or libraries that you need to carry out the project and explain how these will enable the research to be achieved

The code uses novel parallel algorithms and numerical techniques to achieve high performance and accuracy on platforms containing a large number of processors. Desmond supports algorithms typically used to perform fast and accurate molecular dynamics. Long-range electrostatic energy and forces are calculated using particle-mesh-based Ewald techniques. Constraints, which are enforced using a variant of the SHAKE algorithm, allow the time step to be increased. These approaches can be used in combination with time-scale splitting (RESPA-based) integration schemes. Moreover, minimization jobs relax the system into a local energy minimum. The model system is minimized using a hybrid method of the steepest decent and the limited-memory Broyden-FletcherGoldfarb-Shanno (LBFGS) algorithms. The velocity Verlet integration method is used to advance the positions and velocities of the atoms in time and the Particle Mesh Ewald (PME) algorithm will be used to take the full electrostatic interactions into account. There is also FEP/REST algorithm, which enables simulations of a selected subsystem with replicas in a higher effective temperature regime than the remainder of the system, and thus precisely focuses sampling efforts where needed to properly traverse the relevant phase space.

Explain why this project needs the requested VI-SEEM services, why the selected VI-SEEM services are suitable for the project and how the use of the services and resources will enable the science proposed. You should describe the architectural characteristics of VI-SEEM resources that are beneficial to your applications and the problem sizes that have been used to test for scaling and provide supporting evidence (applicable for HPC resources).

The system that we are going to simulate is the FXR bound with known inhibitors. This system consists of ~50000 atoms. The simulation of such large biomolecular systems and the need to accelerate the production of the FEP calculation is computationally demanding. Thus, a significant computation power, as the one that is provided by VI-SEEM facilities supercomputer, is needed. The usage of GPUs is of high importance, since 3 perturbations per day that run on 4 GPUs, would require 768 cores if it was to run on CPUs, thus demanding valuable resources.

Until now we have been running our FEP calculations on Cytera GPU cluster, which makes use of Nvidia M2070 graphics cards. Each FEP calculation has been running on one GPU card and in this way 3 FEP calculations were performed per day. However, in order to be able to run these calculations in 24h, which was the wall time of the cluster, we had to exclude the REST sampling algorithm from the FEP protocol, which enhances the sampling of the calculation. Further more, due to the incompatibility of FEP+ latest versions with the cluster's graphics cards, we had to carry out FEP calculations with an older version of Desmond, not allowing us to yield the benefits of the latest versions. We now intend to use up to 15 GPUs for a FEP calculation, which according to Schrodinger's scaling tests is the most efficient number. In this way, we will be able to incorporate the REST algorithm to our simulations and achieve more accurate results. Apart from that, we will manage to perform our FEP calculations with the latest version of Desmond on your Tesla K40m graphics card. Finally, a significant acceleration of the calculation is expected with newer GPU cards such as the K40m cards, offered by VI-SEEM.

In case you are requesting HPC resources: Justify the number of core hours requested (per code). This should include information such as: run type, wall clock time per step, number of jobs per run type, the number of CPU cores and the total core hours per run type. This information should take the form of a table. Explain how the core hours requested will be used (1 page).

We will use Desmond on GPUs for the following calculations

1) FEP+ of test set compounds for the FXR receptor

10 calculations of 15 compounds each requiring 72 hours x 16 GPUs (10 calc x 72 hours x 16 GPUs = 11,520 GPU hours)

2) Metadynamics calculations of FXR inhibitors

12 calculations of 4 different poses of compounds each requiring 72 hours x 16 GPUs (12 calc x 72 hours x 16 GPUs = 13,824 GPU hours)

Total: 25,344 GPU hours

Describe your experience using resources and services similar to the ones you have requested, in the past and how you will manage using such services and resources. What other experience do you and your team bring to this project?

Dr. Cournia in the period 2003-2006 has extensively used the HEidelberg LInux Cluster System (HELICS) http://i.top500.org/system/5872), which in 2002 was the world's fastest PC-cluster, entering the Top500 at the 35th position. The cluster had 256 commodity of the shelf PCs (Dual AMD Athlon 1,4 GHz, 2 GB RAM, Myrinet NIC M3F-PCI64B-2 (66Mhz/64bit PCI), queuing system: PBS). During this time, Dr. Cournia engaged in large-scale MD simulations of cholesterol and other biological sterols in DPPC model biomembranes with the software CHARMM (scaled up to 64 processors), as well as quantum chemical calculations with the program NWChem (scaled up to 512 processors). Details of this project can be found in (Cournia et al, 2005 and 2007) and at: http://helics.uni-hd.de/users/project_obsolete_overview.php.

In 2013, Dr. Cournia was a PI on the PRACE Project: 2012071260 Project number: RA1260 entitled: “Arp2/3 - Mechanistic studies of the Arp2/3 complex activation”, awarded with 11.2M core hours on CURIE (France). This project ended 4th March 2013. We have used 100% of our allocated time, efficiently managing this Tier-0 system and communicating with CURIE support staff. Since then the Cournia lab has been granted with another three PRACE projects that have finished successfully.

If applicable please explain what type of data you could make available to the VI-SEEM communities if your application gets accepted. What would be the license type? Are there any confidentiality issues with such data?
We will provide the protocols used to enhance the predictivity of structures of ligand-protein complexes. All calculation outputs (trajectories, input + output files) will also be provided.

If applicable please explain what type of software you could make available to the VI-SEEM communities if your application gets accepted. What would be the licence for such software?

If applicable please explain what type of application specific services or workflows you could make available to the VI-SEEM communities if your application gets accepted. What would be the licence for services, workflows?

Group time: Scientific Case
2309.66