SUMMARY STATEMENT

PROGRAM CONTACT: **(Privileged Communication)** ***Release Date:*** **07/18/2019**
Peter Lyster ***Revised Date:***
lysterp@mail.nih.gov

Application Number: **1 P41 GM135452-01**

Principal Investigator

AMARO, ROMMIE E

Applicant Organization: UNIVERSITY OF CALIFORNIA, SAN DIEGO

Review Group: **ZRG1 BST-T (40)**
Center for Scientific Review Special Emphasis Panel
PAR-17-316: Biomedical Technology Research Resource (P41)

Meeting Date: **06/19/2019** ***RFA/PA:*** **PAR17-316**
Council: **OCT 2019** ***PCC:*** **B210LP**
Requested Start: **12/01/2019**

Project Title: **Drug Design Data Resource**

SRG Action: **Impact Score:66**
Next Steps: **Visit https://grants.nih.gov/grants/next_steps.htm**
Human Subjects: **10-No human subjects involved**
Animal Subjects: **10-No live vertebrate animals involved for competing appl.**

Project Year	Direct Costs Requested	Estimated Total Cost
1	1,098,364	1,671,351
2	812,558	1,236,447
3	799,809	1,217,048
4	796,744	1,212,384
5	798,054	1,214,377
TOTAL	4,305,529	6,551,606

ADMINISTRATIVE BUDGET NOTE: The budget shown is the requested budget and has not been adjusted to reflect any recommendations made by reviewers. If an award is planned, the costs will be calculated by Institute grants management staff based on the recommendations outlined below in the COMMITTEE BUDGET RECOMMENDATIONS section.

1P41GM135452-01 Amaro, Rommie

RESUME AND SUMMARY OF DISCUSSION: This application proposes to develop a software platform that will facilitate ligand-protein pose prediction and affinity prediction or ranking that will benefit the computer-aided drug design (CADD) community. The proposed BTRR will utilize the Drug Design Data Resource (D3R) infrastructure and develop three Technology Resource and Development projects. TR&D1 will convert continuous evaluation of ligand pose predictions (CELPP) to a workflow-based challenge that will make container-based workflows accessible to specialists and non-specialists to perform multiple prediction protocols for ligand modeling. TR&D2 will create a compendium of docking, scoring, and ranking workflows for a large number of currently known therapeutic targets or homologous proteins suitable for rational selection by drug discovery teams that will facilitate tool dissemination outside the center. TR&D3 will develop a system to collect data for ligand-receptor docking, affinity prediction/scoring as well as CELPP ranking, workflows, and computational results using the web-accessible public database of measured binding affinities, BindingDB. TR&D3 will also develop a web-based interface and application programming interface for disseminating data to collaborators and the public. In general, the mail reviewers applauded the technical efforts to accelerate small molecule drug docking benchmarking. The Editorial panel, however, thought that the workflow automation and containerization is not the innovation needed to solve current problems with small molecule docking and affinity prediction.

The Editorial reviewers noted that TR&D1 will make CADD methods somewhat more accessible, but that the proposed approach lacked enough targets, method development, and innovation. The reviewers thought the research team lacks docking expertise as well as a discussion of current critical issues in the area of structural based drug design. The lack of method development that will provide docking or scoring improvements was considered a weakness. At the end of discussion, the Editorial panel agree that TR&D1 would facilitate benchmarking, but was unlikely to improve CADD meta-analysis and results, and therefore was likely to have a moderate impact.

The Editorial reviewers viewed the integration of docking protocols in TR&D2 as a logical extension of previous work on affinity predictions which would increase the speed and standardization of analysis. The evaluation of TR&D2 by the Mail reviewers was considered superficial, and the Editorial reviewers thought that while the approach was straightforward it would not have a major impact. The Editorial reviewers commented on the close overlap between TR&D1 and TR&D2 in terms of method and technology development with the main difference appearing to be the type of data. The Editorial reviewers thought that TR&D2 had a relatively superficial connection to the Driving Biological Projects (DBPs), and that the DBPs would not drive technology development. At the end of discussion, the Editorial panel agreed that the chief advantage of TR&D2 would be making the use of the CELPP framework for affinity prediction faster and easier, but would not have a major impact on the field of CADD.

The Editorial reviewers noted that TR&D3 has no unique workflows, is limited to a support role for the other TR&Ds, and doesn't have a substantive independent component. The Editorial reviewers saw no apparent connection or synergy to the DBPs. One of the Editorial reviewers highlighted a concern by Mail reviewer #2 that an overreliance on automatic or superficial curation of the affinity and crystallographic data could become an impediment to the other TR&Ds and DBPs. At the end of discussion, that Editorial panel viewed TR&D3 as a subset of one or both of the first two TR&Ds.

The Editorial reviewers noted that the DBPs focused on software developers and docking methods, and not biology. The Editorial reviewer thought the connection between the DBPs and TR&Ds was superficial, and the feedback the DBPs would provide as highly limited. DBP9 (Drug Discovery @ Home) was considered the one DBP of note in terms of innovation, but that this exercise was done more than a decade ago and doesn't appear to be cited in the application. The Editorial panel did not think the DBPs would drive technology development.

The Editorial reviewers thought Community Engagement plan was strong overall. The D3R is a highly visible resource for docking and scoring, and that software will likely be portable, well documented, and user friendly. One Editorial reviewer expressed a concern that there appears to be two classes of researchers in the proposal – developers and users, and that users don't appear to be engaged, and that Community Engagement was missing a component to address biological questions.

At the end of discussion, the Editorial panel thought the proposal was feasible but that the workflow automation and containerization proposed would not address the current fundamental problems limiting small molecule docking and affinity prediction, and that the impact on the field of CADD would be moderate to low.

DESCRIPTION (provided by applicant): The overall goal of the Drug Design Data Resource (D3R) is to create technologies that will enable dramatic advances in two core purposes of computer-aided drug design (CADD): ligand-protein pose prediction and affinity prediction or ranking. Success in this endeavor will lower the cost and accelerate the discovery of new medications across a range of therapeutic areas. Our approach is based on a recognition that the field currently lacks effective methods to test the accuracy of CADD methods, and that such methods are needed to advance the state of the art. A major challenge is the shortage of unpublished data that can be used to carry out blinded -- and hence objective, large-scale -- and hence statistically significant -- tests. We propose to overcome this challenge by tapping into two large, existing flows of data in a manner that will generate effectively blinded prediction challenges two orders of magnitude larger than are possible today: previously unutilized new PDB co-complexes with druglike ligands, and a flow of newly curated protein-ligand binding data that we will generate using methods already established by BindingDB. Furthermore, CADD methods are increasingly complicated and thus, we need a way to dramatically increase the scale and throughput, as well as the rigor and reproducibility, of the methods employed. We see the development of automated workflows, encapsulating the entire end-to-end CADD experiment, as a key technology to embrace and enable. Thus, we propose to create software and workflow frameworks that will coordinate these continuous, blinded prediction challenges and automatically archive the results, and we will also disseminate and evaluate the results. This work will be done in concert with two classes of Driving Biomedical Project partners. The first class comprises CADD method developers who will use the new frameworks to test and improve their methods. The second class comprises researchers who will apply CADD methods that have been found to perform particularly well in their drug design projects. The latter class includes an innovative plan to crowd-source ligand design, much as FoldIt crowd-sourced protein design. We will engage also extensively with the research community to collect input and maximize the impact of this project.

PUBLIC HEALTH RELEVANCE: Computers are used to help design new drugs, and there is great promise for computational methods to become even more precise and effective than they are today. However, creating improved methods requires giving scientists access to many more experimental measurements of the properties of drugs and drug-like molecules than are now available, so that advanced methods can be tested and optimized. The Drug Design Data Resource seeks to meet this need through the creation of new technologies that will allow researchers more and better access to suitable data, enable systematic testing of their methods, and enhance the ability to share and exchange such methods and their results.

CRITIQUE 1

Significance: 4
Investigator(s): 1
Innovation: 5

Approach: 5
Environment: 1

Overall Impact: This project builds on the substantial experience accumulated by the investigators in four years of running the D3R docking resource coupled with the previous experience in preceding CSAR project that was run at the University of Michigan. The team is excellent and so is the understanding of the issues associated with benchmarking of the docking and scoring methods. It is most certainly important to have ongoing activity focusing on benchmarking molecular docking methods and providing feedback both to the individual scientists and research teams. The proposed approaches to automation and increased ability to handle much higher amounts of docking data are technically sound and the importance of having such resource available to the community is well-articulated. However, it appears that the major novelty is the sheer increase of the number of cases subject to benchmarking and evaluation, a well as, perhaps, providing the infrastructure to individual developers to integrate their tools into the workflow run by D3R. In this reviewer's view, this expectation that the greater access to larger amount of data for benchmarking will dramatically improve respective methods is a questionable supposition. There have been decades of studies many of which have been conducted in private companies and not published but used internally; but thus far, this data has not produced major breakthroughs in docking/scoring outcomes. It would be perhaps more prudent to expect that after about a decade of CSAR/D3R existence, the experience accumulated as a result of method benchmarking would start pointing out some generalized weaknesses or strengths of certain approaches. Unfortunately, such analyses that could drive the investigators' approach appear to be missing. To some extent, the questionable power of this idea of docking quality improvement as a result of greater number crunching was demonstrated by the recent study into ultrafast docking by the Shoichet group (doi: 10.1038/s41586-019-0917-9). Notably, that study did not focus on either of the two major stated objectives of the grant application, ie, predictions of poses and binding affinities. Rather, the authors looked in improved rates of virtual screening, which is a third critical objective of SBDD and perhaps the most desired expectation of medicinal chemists (ie, selecting which molecules should be tested). It is not quite clear why this element of benchmarking the SBDD methods was left out by the applicants.

Another issue that seems to be a missed opportunity was somewhat articulated by the investigators but not developed further. Specifically, they said, that current benchmarking competitions "do not have the statistical power to distinguish clearly between most methods or to let a developer know whether his or her new approach is truly an improvement". The investigators are perhaps in the best position to dive into these challenges by the deep comparative analysis of all submissions but there are no plans that would be also based on past experience that would address this interesting challenge. Finally, the Driving Biological Projects, for the most part, do not pursue important experimental questions. They are technology driving but they are not biological in nature, and the most highlighted project, Drug Design at Home, surprisingly, lacks creative thinking as similar objectives have been pursued elsewhere for a long time. Thus, on the balance the project as described is not overwhelming.

1. Significance

Strengths

- There is always a need to provide benchmarking services to researchers in molecular docking.
- The proposed ability to handle large amounts of docking data addresses the community need as the number of scientists using these methods routinely grows.

Weaknesses

- Unclear if mere increase in the throughput of the Resource along with better tool integration into workflows would indeed dramatically impact method development as no analyses that could identify unique, impactful common elements of either docking methods or scoring functions that would most affect method performance, is intended.

- Poor connectivity to the large community of “simple” users that just need access to one best method to solve their practical design problem and who are confused by the larger diversity of tools

2. Investigator(s)

Strengths

- High level of technical expertise
- Experience with D3R resource
- History of strong collaborations

Weaknesses

- None noted.

3. Innovation

Strengths

- Understanding of the power of end-to-end method integration
- Ability to handle diverse docking packages

Weaknesses

- None of the proposed technical approaches have been developed by the investigators

4. Approach

Strengths

- Organizational structure
- Community engagement; letters of support
- Objective approach to docking results comparison
- Synergy between TR&Ds

Weaknesses

- Although TR&Ds synergize strongly this is also a weakness as they appear more as subsets a single TR&D. One illustration of this point is very high similarity between approaches outlined in Figure 4 on p. 384 (TR&D1) and Figure 1 on p. 446 (TR&D2) that effectively only differ in the input data but otherwise are technologically indistinguishable
- No plans to include virtual screening, which is one of the major uses of SBDD
- The hope that automation and throughout increase would improve methodologies is not well-justified
- No plans to create any new approaches to analyze multiple comparative docking studies and infer unique technical elements that improve docking outcomes
- Majority of DBPs are driving technological rather than biological projects.
- DBP9 (drug discovery @ home) is positioned as one of the most innovative; yet it appears to be ignorant of at least several previous similar attempts both in the US and abroad. For comparison, docking@home (http://docking.cis.udel.edu/) has been tried in 2006-2014; it was an interesting exercise but it did not seem to produce much impact and was closed Another similar project (https://drugdiscoveryathome.com/) with somewhat controversial history has been in existence for at least a decade. So it is unclear what is new and why this new incarnation of an old idea would impact SBDD

5. Environment

Strengths

- Excellent infrastructure is available to support the proposed activities

Weaknesses

- None noted.

Study Timeline:

Strengths

- Timeline is reasonable

Weaknesses

- The expected duration of the resource (10-15 years) is not well justified

Protections for Human Subjects:

Not Applicable (No Human Subjects)

Vertebrate Animals:

Not Applicable (No Vertebrate Animals)

Biohazards:

Not Applicable (No Biohazards)

Resubmission:

Renewal:

Revision:

Applications from Foreign Organizations:

Select Agents:

Resource Sharing Plans:

Acceptable

Authentication of Key Biological and/or Chemical Resources:

Not Applicable (No Relevant Resources)

Budget and Period of Support:

Recommend as Requested

CRITIQUE 2

Significance: 2
Investigator(s): 2
Innovation: 5
Approach: 6
Environment: 1

Overall Impact: The Drug Design Data Resource (D3R) was founded as a U01 Cooperative Agreement, and it is now transitioning to a P41 funding mechanism to support a Biomedical Technology Resource Center (BTRC). D3R became influential by organizing the Grand Challenges that engaged the computational drug discovery community and motivated method development. The goal of the new center to further improve ligand-protein pose prediction and affinity prediction / ranking. The main tools to be developed are CELPP (Continuous Evaluation of Ligand Protein Predictions) for pose prediction and CELPP+ for affinity calculation. The idea of CELPP is to select forthcoming PDB entries that include small ligands, and to release relevant information several days before the structures for testing the docking methods. CELPP also includes a Kubernetes-based server that will host fully-automated, containerized workflows of a number of docking algorithms provided by DBPs. While CELPP will substantially increase the number of docking targets, the main problem is that the automated workflow will incorporate physics based docking methods and cannot fully account for all additional information available on the target protein and the ligand. More generally, although the performance of docking methods heavily depends on the properties of the target, including the flexibility of the protein, the resolution of the x-ray structure, the degrees of freedom of the ligand, the affinity of binding, and very heavily of the fact whether there exist structures with compounds similar to the ligand considered bound to the target, limited resources will be available for target curation. In fact, a preliminary implementation of the approach concludes that very different docking methods yield almost identical overall performance, and hence evaluation on diverse targets yields limited information. It is possible that more information could be extracted by additional curation of the results, but this is barely discussed. CELPP+ will provide high-throughput automation of protein-ligand affinity predictions or rankings. BindingDB curators will collect affinity data and route the identities of the proteins and ligands to CELPP+, a Kubernetes cluster where workflows will predict the affinities. Again, additional curation is the key, and the planned resources are minimal. It is likely that most target affinities will be isolated values rather than carefully curated homologous series from the same laboratory and measured by the same method, the calculated affinities will not be comparable. In addition, employing and even testing methods calculating differences in binding free energy ($\Delta\Delta G$) rather than absolute ΔG values such as FEP and TI will be limited. The co-PIs are experienced computations chemists, but have no first-hand experience in docking, and it seems that the EAB can provide limited help. All innovation in docking and scoring methods is expected to come from the DBPs rather than from the Center. During the D3R challenges the different groups were able to adjust their methodologies for the particular targets, resulting in some development in methodology. However, due to the containerization of the workflows in CELPP and CELPP+ such changes in the methodology, expected entirely to come from the BDPs, will be more difficult to introduce, thereby even hindering method development. Thus, it appears that providing fewer but carefully curated targets in further Grand Challenges would provide more motivation to the DBPs and have better outcome than the planned automated target selection and analysis.

1. Significance

Strengths

- Pose prediction and affinity prediction are among the most important computational steps in drug discovery.

- In spite of the very large number of docking programs, both academic and commercial, it is well recognized the performance of each method heavily depends on the particular application. The success of pose prediction depends on the flexibility of the protein, the resolution of the x-ray structure, the degrees of freedom of the ligand, and the affinity of binding.
- Progress toward improving general pose prediction methods has been slow, and it is unlikely that a general method can be developed that will perform well in all applications. Thus, given a docking problem, the challenge is to select a method that can best use all available information and hence is likely to perform best.
- Similarly, the success of affinity prediction depends on the accuracy of the protein-ligand complex and many other factors, including all information available on the target. Thus, selecting the most appropriate method is a challenge.
- The above description of the current state of docking and scoring indicates the significance of research and need for development toward improved methodology.

Weaknesses

- The big data aspect of the continuous evaluation could be used to judge if a method performs better for a certain protein family. This aspect is important but not elaborated in detail in the proposal.
- It will be valuable if the team could study the sample size needed to reach statistically significant predicting power. However, plans for such – most likely manual – curation of the results is barely discussed.

2. Investigator(s)

Strengths

- The PI, Dr. Rommie Amaro, has excellent educational background and rich experience in computational methods of drug discovery. In particular, she developed methods to incorporate structures from all-atom molecular dynamics simulations in structure-based drug design in a variety of applications, including the investigation of allostery.
- Dr. Amaro has been the director of the National Biomedical Computation Resource, a P41 center with the major goal of developing new multiscale modeling tools and methods to support computational biomedical research.
- The co-PI, Dr. Michael Gilson, is among the most accomplished computational chemists focusing on method development related to receptor-ligand interactions. He is an outstanding theoretician. His contributions includes developing theory and calculation of electrostatic interactions and formulation of models of binding with particular emphasis on host-guest systems.
- Dr. Gilson directs the development of BindingDB, a database of measured protein-ligand binding affinities from the scientific literature and making it searchable and downloadable via a web-accessible database.
- Drs. Amaro and Gilson have been the co-PIs of the NIH U01 project with the goal of establishing the Drug Design Data Resource (D3R), a center aiming at the collection and strategic expansion of datasets from pharmaceutical and academic partners; community engagement and transformation through the D3R challenge activities and workshops. Running the D3R resource clearly provided rich experience to the co-PIs of this P41 application.

Weaknesses

- The Co-PIs and the team of investigators are strong, but have limited hands-on expertise in docking. The data for testing the workflows is likely to require additional curation to determine if such data is appropriate and whether the success or failure in affinity prediction and/or ranking is attributed to the true performance of the algorithms. This expertise seems to be missing in the team.

3. Innovation

Strengths

- The major innovation will be the development of CELPP (Continuous Evaluation of Ligand Protein Predictions) and CELPP+. The idea of CELPP is to select forthcoming PDB entries that include small ligands. Each week, in-house CELPP scripts identify those entries that are suitable for automated docking calculations, including their PDB ID, protein sequence(s), InChIs of any ligands, and the pH of the crystallographic mother liquor. This information will be released several days before the structures themselves for testing the docking methods.
- To test the targets provided by CELPP, the project includes the development of a Kubernetes-based server that will host fully-automated, containerized workflows for pose prediction, created by D3R and DBP (Driving Biological Projects) collaborators.
- Similarly to CELPP, CELPP+ will provide high-throughput automation of protein-ligand affinity predictions or rankings. The idea is that BindingDB curators will collect affinity data and route the identities of the proteins and ligands to CELPP+, a Kubernetes cluster where workflows will predict the affinities.

Weaknesses

- It appears that the only innovative aspect of the project is the development of the CELPP and CELPP+ servers, providing targets and implementing automated docking and affinity prediction methods provided by the DBPs. No potential contributions to the development of methods are discussed. It is possible and is even expected that feedback provided to the DBPs will motivate method development, but no specific plans are discussed.
- While the center will provide automated target collection and evaluation, innovation in docking and scoring methods is expected to come from the DBPs. However, it is possible that due to the containerized workflows the participating DBPs will be more reluctant to adjust the algorithms and thus automation may actually hinder innovation in method development.

4. Approach

Strengths

- It is expected that the development of CELPP and CELPP+ pipelines will increase the number of blinded pose-prediction targets from tens per year to thousands per year, and the number of affinity prediction targets from today's hundreds per year to tens of thousands per year. In principle the large scale calculations provide improved statistics for the comparison of various algorithms.
- The TR&D is associated with 16 DBPs, although a list of eight is highlighted. The TR&D will help DBPs develop deployment containers as part of the evaluation process. The containers enable ready deployment of complex workflows developed by the DBPs. TR&D will provide more objective assessments of the methods developed in the DBPs. Insight from the TR&D has the potential to benefit DBPs by standardizing workflow, data-exchange, and identifying optimal procedures along a complex multi-step workflow for end-to-end execution.
- The container based workflow will enable non-specialist to try out a variety of methodologies.

Weaknesses

- It is stated that the data released from PDB to CELPP will include the PDB ID, protein sequence, InChIs of any ligands, and the pH of the crystallographic mother liquor, all provided several days before the structures themselves are released in PDB. However, CELPP will not be able all available information

- As mentioned, the performance of docking methods heavily depends on the properties of the target, including the flexibility of the protein, the resolution of the x-ray structure, the degrees of freedom of the ligand, and the affinity of binding. In addition, as shown by the D3R Grand Challenges results and noted in the proposal, using existing structural information, such as cocrystal structures of small molecules with the target protein, can increase docking success rates. More generally, all these observations show that there is no pose prediction method that fits all possible targets. The participants of 3DR competition were able to select the best approach for the homologous sets of targets, and it is expected that optimizing such selection algorithms will lead to development in docking methodology. However, the potential for such target-dependent development will be largely eliminated by the randomness of targets coming through the CELPP pipeline.
- The PIs state that the major impediment is inability to compare computational methods in an unbiased, robust, and expedited way. Although this may be true, the quality of the biochemical/biological and X-ray data is even a bigger problem. The platform proposed to be developed as CELPP+ relies on the BindingDB data that contain datasets that may or may not be appropriate for performance evaluation proposed to be hosted by D3R and developed in TRDPs. Additional curation is likely to be needed. Thus, the proposed scale up of the evaluation depends on availability of carefully selected and curated biological data and annotated binding sites. Without these steps, CELPP+ and the DBPs may be swamped with meaningless noise that would prevent learning anything insightful.
- Although it was not demonstrated by the D3R competitions, on a theoretical basis it is expected that most progress in affinity prediction will be provided by free energy perturbation (FEP) and thermodynamic integration (TI) methods that are more rigorous than the semi-empirical and knowledge-based scoring functions. However, FEP and TI calculate differences in binding free energy ($\Delta\Delta G$) rather than absolute ΔG binding free energies. Such calculations can be performed only for homologous series of compounds. Such compounds were provided in the previous D3R competitions, but are unlikely to come through the planned CELPP+ pipeline.
- Based on the above comments, it seems that it would be useful to continue efforts toward collecting pose and affinity data on homologous series as in the previous D3Rs. In fact, the preliminary data on 3,184 blinded ligand-target co-complexes provided by CELPP demonstrate that on the diverse data set all docking methods perform equally well (or equally badly, see Fig 3 on page 70), thus providing essentially no information useful for method development, emphasizing that the real challenge in docking is selecting the best method for a given problem rather than trying to solve all docking problem using the same approach with the same parameters.

5. Environment

Strengths

- UCSD housed the Drug Design Data Resource (D3R), and the environment is excellent to carry out the proposed work.

Weaknesses

- None noted.

Study Timeline:

Protections for Human Subjects:

Not Applicable (No Human Subjects)

Vertebrate Animals:

Not Applicable (No Vertebrate Animals)

Biohazards:

Resubmission:

Renewal:

Revision:

Applications from Foreign Organizations:

Select Agents:

Resource Sharing Plans:

Authentication of Key Biological and/or Chemical Resources:

Budget and Period of Support:

Recommend as Requested

CRITIQUE 3

Significance: 5
Investigator(s): 1
Innovation: 6
Approach: 5
Environment: 1

Overall Impact: Although the Drug Design Data Resource (D3R) is valuable, and the Grand Challenges are helpful, a full-blown P41 Center to support such activities is overkill and unnecessary. Also, the fundamental issue of docking, which is the conformational change induced by the bound ligand is not adequately addressed in the application. Moreover, although many papers have been generated through the past Grand Challenges, there is no clear evidence to indicate that what had been learned from those studies have been implemented into the current docking technologies.

1. **Significance:**

2. **Investigator(s):**

3. **Innovation:**

4. Approach:

5. Environment:

Study Timeline:

Protections for Human Subjects:
Not Applicable (No Human Subjects)

Vertebrate Animals:
Not Applicable (No Vertebrate Animals)

Biohazards:

Resubmission:

Renewal:

Revision:

Applications from Foreign Organizations:

Select Agents:

Resource Sharing Plans:

Authentication of Key Biological and/or Chemical Resources:

Budget and Period of Support:
Recommend as Requested

TECHNOLOGY RESEARCH AND DEVELOPMENT PROJECTS (TR&D)

TR&D 1-Drug Design Data Resource: Pose Prediction

Priority Score: 56

CRITIQUE 1

Top Score Drivers:

- The D3R team has extensive expertise in conducting blind challenges of protein-ligand docking. The team is highly visible in the field and can exert large influence on the future of the field of computer-aided-drug-design.
- The proposed high throughput automated testing protocol is more powerful than what is currently available. The insight of high throughput testing potentially will benefit the development and refinement of methods for protein-ligand binding.
- The container based workflow to be developed with the help of a P41 would facilitate the dissemination of the prediction toolchain provided by the DBPs. At the same time, the containers enable easy deployment of the toolchain on cloud based computational resources.

Quality of the Research:

Although the TR&D in itself does not directly address the development of better methods for protein-ligand binding, it is valuable in that it provides a more robust evaluation of existing protocols. The team plans to help research groups convert their protocols into automated workflows that runs in containers for easy deployment.

Preliminary results (CELPP challenge) are good but it showed no clear improvement of the method evaluations over the smaller scale Grand Challenge conducted previously by the D3R team. The bottleneck for ligand design is the development of the actual models for pose prediction and affinity determination. Model evaluation being proposed is of secondary importance but still indispensable.

With container based deployment only has access to a local PBD, one would image it is not necessary to use prerelease PBD co-crystals. A local copy of pre-2019 PBD could be made available to the workflow, with the evaluation preformed on 2019 targets. One could also randomly pick 1000 structures as a test set and only make the remaining structures in the PBD available to the workflow.

The big data aspect of the continuous evaluation could be used to judge if a method performs better for a certain protein family . This aspect is important but not elaborated in detail in the proposal. It will be valuable if the team could study the sample size needed to reach statistically significant predicting power. More ambitious evaluation guided protocol refinement will likely need a larger sample size.

Although no alternative approaches are presented, the technology deployment strategy seems to be rather straightforward; borrowing mature technology from a different field (information technology) to biological modeling problems, the likelihood of success is high.

Minor quibble aside, the TR&D represents state of the art model evaluation and deployment technology in an important area of research and will likely contribute to advance the frontiers.

Overall Technology Development Program:

The pose and affinity TR&D projects are complementary. The model evaluation TR&Ds directly serves several highly successful DBPs. Tools developed by the DBPs are used by a large community of scientists. Through close collaboration with the DBPs, a synergistic push will be made to set standards and best practices for protein-ligand binding.

There is indeed a coherent vision in implementing the advanced evaluation techniques for pose prediction and affinity determination.

The entire arc of the technology development process is projected to be 15 years with the first 5 years being framework development and DBPs recruitments. In the second 5 years, the TR&D is expected to provide more substantial service to the DBPs. DBPs may start to deploy in-house containers developed in collaboration with the TR&D team. The final 5 year goal is not fully defined at least as part of the overall research strategy due to anticipated changes in the field.

The pose prediction related TR&D (TR&D1) is expected to be carried out to completion. I see no major technological impasse. However, it is possible the actual pose-prediction algorithms won’t advance far

enough to have a stand-out winner, leading to ambiguous evaluations if all major players score similarly. Insights from the evaluation process may help advancing the pose-prediction algorithms.

The TR&D mainly serves the community of model developers for pose prediction and affinity determinations. It seems straightforward to make the container based workflow available to non-specialists, thus enabling scientists who are consumers of advanced ligand modeling methodologies to try out multiple prediction protocols easily.

Insights from the evaluations conducted as part of the TR&D projects will benefit DBPs. More extensive evaluations of the methods by the TR&D will guide teams working on different pieces of drug discovery and lead optimization to identify the best approaches for other pieces needed for their work.

The TR&D provides new capacities not currently available. Pose-prediction and affinity determination are of fundamental significance for biomedical research. That being said, the TR&D only addresses the evaluation part of the pose-prediction and affinity determination. Although evaluation is indispensable, it has to be coupled with advances the actual pose and affinity predictions (DBPs) to make the biomedical impact.

The TR&D personnel is suitably qualified to lead the projects.

Technology Development Partnerships:

There is no technology development partnerships, but the data sharing through the extensive industrial partnership is big plus for the TR&D.

Renewal Applications (if applicable):

DBP(s) associated with this TR&D: all DBPs.

Overall: The TR&D is associated with 16 DBPs according to the table provided, although a list of eight is highlighted. Specific interactions with each individual DBP are not fully discussed making assessment of interactions more challenging.

The TR&D will help DBPs develop deployment containers as part of the evaluation process. The containers enable ready deployment of complex workflows developed by the DBPs. TR&D will provide more objective assessments of the methods developed in the DBPs. Insight from the TR&D has the potential to benefit DBPs by standardizing workflow, data-exchange, and identifying optimal procedures along a complex multi-step workflow for end-to-end execution.

Although a mutual beneficial relationship is expected, specific examples of the interactions are somewhat lacking in the portfolio descriptions of the DBPs. The DBPs are expected to advance and motivate the TR&Ds in that specific evaluation needs of the DBPs will likely shape the TR&D. However, the DBPs and TR&D are not intimately intertwined in that not pursuing the TR&D may not have large negative impact on the DBPs.

DBP/TR&D Interactions:

There is a synergy between TR&D and DBPs in that the TR&D providing critical evaluation and container based deployment vehicles for the DBPs. The DBPs drives the optimal modularize development of the TR&D workflow and data exchange. Potentially DBPs could benefit from insight gained from the evaluations.

Renewal Applications (if applicable):

CRITIQUE 2

Top Score Drivers:

1.Collecting different methods. 2. Collecting different data. 3. Integration.

- There are several distinct goals of the proposed approach which are crucial for lead compound discovery
- Convert continuous evaluation of ligand pose predictions to a workflow-based challenge
- Develop and deploy an advanced website to share convert continuous evaluation of ligand pose prediction results and workflows
- Analyze CELPP continuous evaluation of ligand pose prediction results to extract scientific insights and value
- There is an issue about the completeness of the method and data collection in this proposal. Some methods and data resources seem missing.

Quality of the Research:

The TR&D proposed is updated frequently from different methods and results from different sources. The drug pose prediction is an important research area, critical for drug lead compound discovery. The TR&D developed should be used to advance the frontiers of the biomedical research for drug compound discovery. Alternative approaches were suggested to solve the technical issues which may be encountered.

Overall Technology Development Program:

This project is synergistic with a coherent vision for BTRR. The core idea and procedure were addressed in the project description. It was stated that the technology development can be carried out through the completion with the optimization. Efforts should still be made further so that non-expert can use. This proposed study will help the biomedical research for more accurate and reliable molecular level drug discovery. The Resource will be useful for the general community and perhaps the hospitals as well as the doctors for making use of the results from the cutting edge research for the drug discovery. The resource technology is only partly available. Some need more developments and others needed to be pieced together. The Resource TR&D personnel is experienced and qualified to lead these proposed projects.

Technology Development Partnerships:

Technology development partnerships were proposed and justified. It appears that the planned partnerships can lead to successful integration of partner technologies into the resource.

Renewal Applications (if applicable):

DBP(s) associated with this TR&D:

Overall: The significant technology applications appear to be based on the different methods developed and data collected. The proposed DBP should be served as the test bed for the TR&D research project specific to the pose prediction. DPB may motivate R&D in the resource. The technology for proposed DBP may have impact on science. The described interplay between DBPs and TR&Ds in the proposal seems reasonable.

DBP/TR&D Interactions:

There is a tie and synergy between the TR&D project and the DBP in advancing the focal technology.

Renewal Applications (if applicable):

TR&D 2- Drug Design Data Resource: Ligand Ranking & Affinity Predictions

Priority Score: 57

CRITIQUE 1

Top Score Drivers:

- The plan for the development of the platform to compare affinity prediction and ranking (CELPP+) is clear and well-thought through. A web-based demonstration is present and the overall approach has been tested.
- The integration of TRDP2 with other TRDPs is natural and essential for solving the overall biomedical problem – development of robust methods for prediction of small molecule-protein interactions. The format and the involvement of the partners in DBPs would allow a robust testing of the platform. It is synergistic in nature and would accelerate exchange of knowledge between DBPs in an unbiased and timely fashion.
- The team of investigators is strong, appropriate for the efforts proposed, yet it has a moderate weakness. The ultimate goal of the project is facilitation of drug discovery projects. However, the team of investigators has a very limited hands-on expertise in this area. The data for testing the workflows is likely to require additional curation to determine if such data is appropriate and whether the success or failure in affinity prediction and/or ranking is attributed to the true performance of the algorithms. Generation of insights in aim 4 require a set of experts with drug discovery. This expertise in missing in the team.

Quality of the Research:

The TRDP 2 relies on the currently implemented and tested CELPP/D3R Grand Challenges infrastructure, scientific approaches, and technical implementation. These efforts facilitate search and development of the most appropriate CADD methods for accurate methods for prediction of small molecule - protein affinity and relative ranking, key steps in modern drug discovery projects. The search for such methods is usually tedious, slow, and consists of multiple iterations, manual and, often researcher-biased, selection of the docking, scoring, and ranking methodologies appropriate for the therapeutic target of interest.

Although similar approaches and technical implementations exist and have been listed in the proposal, these technologies/resources/efforts are too small in the scale/magnitude and the scope compared to those proposed in this application. The comparison of the docking/virtual screening methods implemented in the PI’s previous D3R Grand Challenges and those by other research groups, although useful, are too small in size to show statistically significant difference in the accuracy of the methods. Hence, efforts proposed in this TRDP are fully justified.

The proposed efforts will generate a compendium of docking, scoring, and ranking workflows for a very large number of currently known therapeutic targets or homologous proteins suitable for rational selection by individual drug discovery teams and quick implementation on the local computational resources. Considering tight integration with BindingDB and many other publicly accessible databases, this effort should be able to serve not only as a sandbox for testing the novel state of the art

computational approaches to CADD (developed in the DBP) but also to disseminate the findings and to connect the CADD developers with the drug discovery teams that would benefit from these new technologies.

Alternative approaches for solving this particular technological problem have not been presented. Since the working model of the technology already exists and have been tested by the PI, DBP partners, and community, it is unclear if the alternative approaches are needed for the software/hardware part of the platform that would test the affinity prediction and ranking algorithms. However, addressing the extraction of the insights process in aim 4 may require multiple alternative approaches to be tested, which are not presented. For instance, affinity prediction and ranking are highly context dependent (open or closed binding site, solvent exposed, flexibility/induced fit, etc), and the appropriate methods are either developed or the appropriate methods are chosen by the researchers who handle this specific problem. It is unclear if this information will be passed on to the CELPP+ workflows from the PDB or if the BindingDB would extract and contain this type of data during the X-ray model import step from the PDB.

The PI states that the major impediment is inability to compare computational methods in an unbiased, robust, and expedited way. Although this may be true, the quality of the biochemical/biological and X-ray data is even a bigger problem. The platform proposed to be developed relies on the BindingDB data that contain datasets that may or may not be appropriate for performance evaluation proposed to be hosted by D3R and developed in TRDPs. Additional curation is likely to be needed. It is mentioned that the appropriate error metrics will be defined together with the CADD community, but it does not appear to be sufficient. The success of the evaluation step in aim 4 hinges on yet to defined criteria and no alternative approaches are proposed. It seems that the proposed scale up of the CELPP+ project 2 depends on availability of carefully selected and curated biological data and annotated binding sites. Without these steps, the BindingDB and the DBPs may be swamped with meaningless noise that would prevent learning anything insightful.

Overall Technology Development Program:

All the proposed TRDP projects are highly complementary and synergistic as they are essential components of the task typically addressed by pose and affinity prediction and ranking (TRDP 1 and 3). The overall vision of the proposed TRDP program is coherent, and is largely based on the overarching hypothesis that further CADD method development and implementation is impeded by the lack of a standardized technological platform for robust readily interpretable comparison of the existing and newly developed methodologies for docking, scoring, and ranking as applied to drug discovery projects. The proposal has a plan to integrate the workflows and the results of the data to the already widely used BindingDB database publicly available on the web. The plan also includes testing the technological solution with two researchers whose primary research interests are development of novel small molecule therapeutic agents, Drs Harki and Wagner, who are considered non-specialists as it pertains to the development and implementation of the proposed technology. If successful, the proposed TRDP will synergize efforts of multiple biomedical researcher groups that are involved in drug discovery either directly or indirectly. The proposed resource is unique in the technological objectives it is trying to achieve, the scope of the involved DBP components, the magnitude of the effort, and integration with other already resources. The resource is already available in its limited form and generally utilized by multiple academic and industrial researchers. The proposed resource is also unique as it would allow to combine a diverse set of affinity and ranking predictions evaluated in a standardized fashion across multiple protein targets and small molecule chemotypes, resulting in an unbiased and independent comparison of the methods, parameters, datasets, etc.

The technology development in TRDP2 is likely be carried through to completion as the PI and the team has committed to the project as evidenced by successful implementation of the previous iteration of CELPP+/D3R. The PI also has secured expertise and interest of the method development researchers and the non-specialists listed in the DBP, who will be the users of the platform. My only concern is that the team has no internal expertise in all the aspects of Aim 4 (see below). This may

potentially lead to the situation where the usefulness to the drug discovery efforts, which is the ultimate goal of the resource, is minimal.

The personnel involved is highly qualified to lead these projects. A demonstration of their qualifications already exists as evidenced by CELPP/D3R platform and multiple letters of support. This is a strength. That said, the team of researchers is largely CADD oriented with expertise in multiple areas of computer-aided/computational biomedical research. Many aspects of the efforts proposed rely on curation and selection of the data that will be used to test the pose and affinity prediction and ranking as applied to drug discovery. None of the team seems to have hands-on experience in drug discovery, which is a moderate weakness of this proposal. Linda Hwang is listed as a person responsible for curation of the data. Her qualifications and expertise are unclear since her biosketch is not present. Although BindingDB can provide data that have been already curated, it cannot provide high level of scientific insights into the quality of the assays and the experimental data necessary for analysis of the outcomes of docking. It is unlikely that partnerships with Drs Harki and Wagner are sufficient to cover these aspects. It is unclear who will be responsible for extracting the insights in aim 4. To be able to do all this, the team would need to have a vast expertise not only in computational methodologies but also in modern screening techniques, assay development, medicinal chemistry, biology, etc. None of this expertise is present in the team. It does not appear that the EAB would be able to contribute to this effort either. Although community involvement may provide the necessary evaluation of the docking efficiencies and underlying biological data, this effort may quickly degrade if the quality of the data used for optimization of the workflow is poor.

Technology Development Partnerships:

Not applicable

Renewal Applications (if applicable):

DBP(s) associated with this TR&D: Drug Design Data Resource: Driving Biomedical Projects

Overall: It appears that the PI and the team vision for DBPs (only one is included in this application) relies primarily on developing a platform for the partners who contribute their computational methodologies to the workflows, design and evaluation of interaction with BindingDB and D3R data uploaded to BindingDB, and testing the user interface. It is anticipated that availability of the platform and the standardized way to interact with the platform will facilitate development of the better docking, scoring, and ranking methods.

These partnerships include 14 different groups of researchers contributing their codes to the technology development. Importantly, it is envisioned that by participating in TRDPs and DBP, the partners will have access to quick (weekly) and unbiased evaluation of the performance of their methods and also obtain insights into their own code optimization and improvement of the software overall. There are also two partners Drs. Drs Harki and Wagner who will test the practical application of the technology platform proposed to be developed. A plan for collaborating with non-specialists is outlined and appears to be limited to user interface testing and evaluation of the decision-making at the point of the end-user. Although it may be sufficient, the team may benefit from additional testing in real-life drug discovery projects to demonstrate the power of the resource with a wider variety of the biological targets and difficulty of the projects.

DBP/TR&D Interactions:

The proposed DBP interactions will include multiple method and software developers from academic and industrial laboratories. This is appropriate and desirable to ensure that standardization of the technological solution accounts for a large variety of possible permutations of the required input data, complexity of algorithms and the required computational resources. A description of the partnership

with Drs. Drs Harki and Wagner, non-specialists in the development of technology, is present and a rationale for their involvement is outlined. The technology development partners are committed to participating in TRDP2 efforts as evidenced by their letters of support.

TRDP2 and DBP are the essential components of the proposed resource. It is anticipated that both TRDP2 and DPB would benefit from gaining access to a unified platform to test affinity prediction and ranking methods by rapid testing of the technologies with the same datasets. It is also anticipated that a direct comparison of the affinity prediction and ranking methods would identify common strength and weaknesses, hence, it would facilitate better understanding of the underlying scientific problems.

There is clear synergy between TRDP2 and DPB participants in advancing the focal technology within the proposed resource as well as in advancing the underlying science of the DBP participants.

Renewal Applications (if applicable):

CRITIQUE 2

Top Score Drivers:

- The pressing need for a rigorous approach for comparing and validating methods for predicting protein-ligand binding affinity
- Creation of modular workflows instantiated in current state-of-the-art container framework
- The need to expand the community of partners to include more of the leaders in affinity prediction, especially those developing and applying free energy perturbation and related methodologies

Quality of the Research:

Computer-aided drug design, while widely used for impact in the pharmaceutical industry, has reached a plateau, with few true innovations or improvements in methodology. An important root cause of this stagnation is the inability to rigorously compare and validate methods across a wide range of test cases, especially test cases that are blinded. TR&D 002 therefore is likely to advance the frontiers of biomedical research by providing the required rigorous assessment methodology and thereby enabling a fresh burst of methodological improvements. The work proposed here would have value in its own right in the development of methodologies for statistically relevant comparisons and the identification of important factors leading to success or failure of affinity prediction methods.

Overall Technology Development Program:

TR&D 002 will provide a critical resource to the computational research community, for the first time enabling a rigorous approach to validation of methods for binding affinity prediction. There is currently no comprehensive way to assess these complicated, multi-parameter methodologies, and therefore the published literature provides only a collection of anecdotal application of these methods to protein-ligand affinity data sets. The resource to be created here offers the potential for a systematic understanding of the current state of the art of these methods, allowing careful assessment of the factors influencing performance. TR&D 002 is well integrated into the other two TR&D projects, which strengthens the likelihood of success for TR&D 002. The co-PIs have already containerized two docking codes within the TR&D 002 framework, which provides some evidence of their ability to containerize affinity prediction codes at a higher level of theory than docking; it is likely, however, that the team will face challenges in early years in how best to containerize and document the more complicated, more parameter-rich affinity prediction methodologies.

There is a high likelihood that the technology development will be carried out to completion; the primary leading indicator of full completion will be the successful transition from the team’s current

containerized workflows of docking for affinity prediction to a demonstration of at least one fully containerized workflow for a more complex affinity prediction method such as free energy perturbation calculations. The TR&D 002 team shows evidence of deep expertise in both the scientific aspects of the computational methods and in the instantiation of suitable computational infrastructure and are well placed to lead this project. Close partnership with DBPs 6 and 14 in particular will give the team direct feedback on how best to implement these workflows for use by non-computational, non-specialist users.

Technology Development Partnerships:

The co-PIs have assembled a strong set of partners spanning docking codes, physics-based and machine-learning affinity prediction methods, and application to real-world drug discovery problems. Partnering with the MolSSI and BioSimSpace initiatives is an important strategy for moving these methods towards robust, modern computer science practice. However, this TR&D would be significantly strengthened if the list of partners included some of the more prominent research groups who develop and apply affinity prediction methods. The co-PIs have assembled a subset of their DBPs that focus on affinity prediction at higher levels of theory than docking; those researchers are solid, but it would be good to expand to include one or two more of the key influencers in the field who could encourage active participation in containerizing affinity prediction workflows.

Renewal Applications (if applicable):

DBP(s) associated with this TR&D:

Overall: The crowd-sourced DBPs, especially DBP 9 which includes Boehringer Ingelheim, are likely to present interesting drug discovery challenges for the workflows being developed; these would be expected to involve protein systems beyond the types (e.g., protein kinases) typically seen outside of the pharmaceutical industry and will provide a useful test bed for both pose prediction and affinity prediction workflows. In addition, the continuation of D3R Grand Challenges is likely to incubate a flow of new DBP relationships, and the containerized workflows to be developed will enable other researchers to make use of affinity prediction methodologies without necessarily becoming official DBPs.

DBP/TR&D Interactions:

While the full list of DBPs has the strongest connections to TR&D 001 due to the history of D3R, DBPs 4, 7, 8, 9, and 10 in particular will be instrumental in building out the initial containerized workflows for TR&D 002 and will allow for a rich interplay between validating current state and iteratively improving future state of affinity prediction methods.

Renewal Applications (if applicable):

TR&D 3 - Drug Design Data Resource: Data & Data Analytics

Priority Score: 59

CRITIQUE 1

Top Score Drivers:

- Several proposed advance on data analytics.
- Collecting and curating the protein-ligand interaction data.
- Archiving workflows and challenge results with associated protein-ligand interaction data.
- Serving data to project websites, CELPP+ workflows and the research community

Quality of the Research:

TR&D continues updating. Supposed to provide the newest data collection of the protein-ligand interaction data. This is useful for the biomedical research especially the drug lead compound identification. Alternatives have been proposed.

Overall Technology Development Program:

The proposed TR&D projects are synergistic with a coherent vision. The main procedures of the technology development process were described in the TR&D projects. The technology development and data collection appear to be carried through to completion with optimization. Improvements are still needed for use by nonexperts. This will help to serve a platform for biomedical researchers to use in identifying potential drug targets. This data resource will be convenient and useful to the research community and for further advances in research on drug discovery. The data resource technology is only partly available and will need to be developed further. The data resource TR&D personnel are qualified to lead these specialized projects.

Technology Development Partnerships:

The technology development partnerships were proposed and justified? It seems that this will lead to a successful integration of partner technologies into the Resource.

Renewal Applications (if applicable):

DBP(s) associated with this TR&D:

Overall: The proposal for developing significant applications is based on the high-quality data collection from biomedical research. The proposed DBP can be used as a driver for speeding up the data collection and platform construction. The DBP study will certainly motivate technological R&D in the Resource. The technology for software development, serving data to project websites, CELPP+ workflows and the research community is suitable for the proposed DBPs and is expected to have impact on the science studies. The proposed DBPs will drive for TR&D research. The described interplay between DBPs and TR&Ds is useful.

DBP/TR&D Interactions:

There is a tie between the TR&D project and the DBP in advancing the technology.

Renewal Applications (if applicable):

CRITIQUE 2

Top Score Drivers:

- The overall plan for the improvement and development of novel capabilities within BindingDB to acquire ligand-protein affinity data, co-crystal data, and dissemination of the achievements is

clear and has been previously tested with a similar type of data. The BindingDB project is well respected and serves as a free and easily accessible source of information for drug discovery projects. That said, overreliance on automatic or somewhat superficial curation of the affinity and crystallographic data may result in the overall impediment of the technology development in TRDP1 and 2 and DBP.

- TRDP3 and other parts of the proposal are complementary and synergistic and are likely to succeed if the quality of the data supplied to TRDP1 and 2 and DBP is high.
- The team of investigators is strong and has exceptional track record and demonstrated example of working database solution (BindingDB). The team may benefit by including personnel that will be able to guide, implement, and execute additional necessary curation efforts as well as experts in assay development, crystallography, and other aspects of drug discovery.

Quality of the Research:

In this proposal, the PI aims to develop a technology solution for collecting data for ligand-receptor docking, affinity prediction/scoring, and ranking and workflows and computational results (CELPP+ results) in the BindingDB as well as development of web-based interface and API for dissemination of the data to the collaborators and the public. BindingDB is one of the two databases in the world (ChEMBL is another one) that contains comparable data relevant to development of small molecules as therapeutics. Both BindingDB and ChEMBL exchange their data, which leads to a faster overall growth of the curated data in both databases. Addition of curated crystallographic data, docking workflows, and outcomes of the workflow comparisons to BindingDB would further contribute to increasing its value to laboratories working on drug discovery as well as to broader research community.

The overall quality of research efforts in aims 2 and 3 is high and appropriate for advancement drug discovery to new frontiers.

Aim 1 may require alternative approaches. The way it is written indicates that the curation of the data will be done in an automated or semi-automated fashion. Although the proposed minimal curation (based on the proposed personnel effort) this may work in very straightforward cases (e.g. ATP binding site in kinases), anyone who worked on a drug discovery project knows that it is a much more involved process. It is stated that 50,000 (+25,000) data (which seems to mean activity/affinity for one compound) will be curated per year. This would translate to 1000 sources (50 compounds per publication, which is a very optimistic estimate) curated automatically or with the help of one FTE working ca 2-3 month (50-75 days). This is an overly ambitious goal. It is difficult to imagine that high quality curation is possible using the process described in aim 1. Unless the protein target is well explored, and large amount of verifiable data are generated, which would also mean that the target is unlikely to be of interest to non-experts, the “manual” curation of one target and its ligands would require at least one day by a highly skilled expert.

The same concern is true for collecting and curating crystallographic data. Since the success of the resource depends on the quality of the biological and X-ray data used to test and optimize workflows, absence of a sound and convincing to drug discovery experts plan for data curation is disappointing.

Overall Technology Development Program:

TRDP3 is an essential cheminformatics component of TRDP1 and 2. It is both complementary and synergistic. None of the efforts proposed in TRDP1 and 2 are possible without TRDP3. On the other hand, further increase in the number and variety of the datasets in TRDP3 available for TRDP1 and 2 would accelerate development of better workflows and overall success of CELPP+. The description of TRDP3 is well thought through (except aim 1), have well described plan, and includes a working demonstration of what to expect since BindingDB is functional and available for use by the public worldwide. The proposal covers all the necessary aspects of technology development process except high quality curation of data. BindingDB is already a unique compendium of drug discovery relevant

data. Additional inclusion of the X-ray and docking data would further strengthen its utility both as a standalone resource as well as to serve the input data to the TRDP1 and 2.

Several aspects of data collection appear to rely primarily on the existing approaches, community effort, and possible involvement of the members of EAB. The team itself does not appear to have sufficient resources and expertise to assess the quality and usefulness of the data. The automated collection of the affinity/activity data without careful curation would not result in the outcomes expected (see also comments in the previous section). It is proposed to use an automated process to link the affinity/activity and X-ray data. Absence of sensible curation before using X-ray data may result in unpredictable results. For instance, it is not unusual to find multiple copies of the protein and small molecule ligands in co-crystal structures. Sometimes these multiple copies are just what they – multiple copies in the asymmetric unit cell, sometimes they are a part of the dimer, trimer, tetramer, etc, and sometimes they form completely artificial multimer interaction found only in the X-ray structure. Curation of X-ray structures for novel or less explored targets is not an easy task and may take one day of intense investigation of the topic by an advanced PhD level highly skilled expert. It does not appear that such level of curation is either proposed or even possible with the personnel involved. Inclusion of the ligand-protein complex without rigorous curation, on the other hand, may result in an increase in the overall noise and distortion of the “insights” developed in TRDP1 and 2. The quality of data – affinity, co-crystal – is the cornerstone of the success of the proposed technology. This aspect is almost completely ignored in the proposal.

As stated, the overall goal of the proposal/resource is to enable dramatic advances in development of novel CADD methods, algorithms, approaches. Accumulation of high quality data proposed in Aim 1 does not appear to be possible with the methods proposed. Since availability of the high quality data is key to advancing to the next level in CADD, completion of the project and use of the resource by non-specialists may be impeded.

The TRDP3 (and other TRDPs) personnel is suitable for the majority of the proposed activities. However, it is difficult to imaging how dramatic advances in CADD are possible without active participation of experts in drug discovery other than those involved in CADD, i.e. biologists, crystallographers, medicinal chemists, etc, and appropriate support and guidance for collection and appropriate curation of affinity/activity and X-ray data.

Technology Development Partnerships:

Renewal Applications (if applicable):

DBP(s) associated with this TR&D: Drug Design Data Resource: Driving Biomedical Projects

Overall: The improvement in the ligand placement, affinity prediction, and ranking methods is the major function of the DBP projects. It is expected that the DBP partners will develop and upload their workflows for testing using the unbiased methods developed by TRDP. There will be a constant interaction between DBPs partners and TRDPs to improve and refine the technology for data storage and exchange of the performance data. The two non-specialist experts Drs Harki and Wagner will test the user interfaces and usability of the data uploaded by the workflows for their research.

DBP/TR&D Interactions:

The proposed DBP partnerships are an essential part of the technology proposed to be developed. The technology is fully dependent on the development of workflows and testing by the the DBP partners and possibly by a wider research community. There is clear synergy between DBP and TRDP.

Renewal Applications (if applicable):

DRIVING BIOMEDICAL PROJECTS:

Priority Score: 70

CRITIQUE: The issue of this DBP is that the PI and other co-investigators of the project fail to demonstrate the scientific significance of the proposed studies. There are no preliminary data to show why the proposed DBP is scientific significant. For example, for each docking, only one pose, the correct one, is useful; however, in the application, the reason for the evaluation of other poses is not very clear. There are a lot of statements in the proposal; however, there are few data to back up those statements.

Renewal Applications (if applicable):

COMMUNITY ENGAGEMENT:

Priority Score: 32

CRITIQUE 1

Quality of the Community Engagement Plans:

The Drug Design Data Resource (D3R) was founded as a U01 Cooperative Agreement, and is now transitioning to a P41 funding mechanism to support a Biomedical Technology Resource Center (BTRC). The Center has a cogent and tightly focused goal concerning a particular need of the research community, namely the necessity to rationally improve computationally-driven drug design (CADD) methods in terms of precision, effectiveness and throughput in the two key challenges of CADD, ligand-protein pose prediction and affinity prediction / ranking, with the ultimate goal of creating new and improved medications.

Because the overarching goal of the D3R is to address a pressing need in the biomedical community for effective methods to test the accuracy of computer aided drug design methods and increase their size and scope, an effective Community Engagement strategy is of particular importance for this Center (as stated, “community engagement is at the heart of D3R”). The Center has already established an extremely strong track record in this regard, and as noted the Center’s success has been predicated on intense community-centered efforts, making D3R an important hub for CADD developers throughout the community; indeed, there are already potentially over 12,000 registered users through the online resources such as HADDOCK (Utrecht) and AutoDock Vina (Scripps) already available through Center members and collaborators.

A multifaceted, well integrated and already proven Community Engagement tactic is presented by a highly qualified team in order enhance the Center’s visibility among researchers, including the community challenges (Aim 1), organizing workshops, “hackathons” and online training activities (of which the Biomedical Big Data Training Collaborative online resource and the innovative webinars, with excellent attendances, are already excellent examples) (Aim 2), additional expansion of web-based outreach (Aim 3), presentation / hosting at conferences, and regular publication (also called Aim 3 but presumably Aim 4). The Center plans on strong leveraging of available resources in the community to further expand their outreach. In this, it is particularly important how the Center has already engaged the pharmaceutical and biotechnology industries (sometimes overlooked by other BTRRs) in companies such as Novartis and Janssen, in particular to provide and collaborate on their extremely valuable protein-ligand datasets – and numerous data sharing agreements to supply large amounts of

unpublished ligand-protein interaction data are already in place. Innovation is mainly focused on the improvement of many of the most proven successful approaches and expansion of users, as for example in the implementation of new “rolling challenges” to complement the existing Grand Challenges, and extending training sessions to webinars with significantly greater outreach potential. Collectively it is clear that the D3R’s Community Engagement plans are practical, impactful, sustainable and can reach and can be readily accessed a broad range of relevant biomedical researchers.
It is particularly notable that there is a well-conceived plan for how community engagement will be pursued and matured over the projected 10-15 year arc for the Center. The Center starts with the strong advantage of building upon already-initiated and successfully implemented means for interaction with the community, such as the annual Grand Challenges and workshops, and much of the initial work will focus on expanding upon the existing software and administrative / organizational infrastructure to support the community engagement, with a gradual switch to individual research grant support, automation (and so reduced costs), commercialization and in-house installation as the Center matures, but while maintaining strong archival plans using recognized and durable public databases.

Community Engagement Training:

The training sessions and webinars are yet another example of how the Center has already implemented highly successful facets of its overall Community Engagement strategy. Starting with a suitably qualified and highly experienced team, there is an outstandingly logical attention to detail of how to take resources and developments at the Center and implement them as training, by starting with the annual workshops and developing from there, and for example recognizing that the modularization of workflows naturally lends themselves to division of corresponding training modules. The Center is also already keeping track of various metrics to assess the success of their outreach efforts including online user numbers, user workloads, and workshop attendances (e.g., 55 participants in the Grand Challenge 4, the metrics of CE Fig 5, location data as in CE Fig 4). While no formal collaboration service component is proposed during the first phase of the D3R as a BTRR, however, in the context of the TR&Ds, DBPs and other strong outreach efforts such as training and challenges, the proposed model seems very appropriate and seems extremely likely to continue in its already very successful efforts to enable adoption of the D3R’s resources and approaches in the wider community.

Dissemination of Resources:
The Center presents an ambitious, extensive, yet practical and proven approach to rapid and widespread resource dissemination, which involves a great deal of productive two-way engagement with users, developers and other members of the biomedical community, involving efforts already discussed above. The ultimate metric of success, already being effectively monitored as a substantial part of the Center’s efforts, is the significant improvement in the accuracy of protein-ligand pose and affinity predictions due to D3R activities. The D3R thus distinguishes itself by not considering classical publication as the major metric, rather it is how they can improve the community that is their crucial direct metric. The Center for example has the laudable objective of expanding the current approach of holding annual blinded prediction challenges, that currently use only tens of crystal structures and several hundred affinities as test beds (which has proven insufficient to clearly draw statistically rigorous conclusions concerning which CADD methodologies are performing better in particular areas). This expansion is envisioned to involve the development of an integrated technology set to assess CADD methods with thousands of test cases per year. This will be accomplished through the generation and widespread dissemination of effective methods and community-wide blinded prediction challenges to test the accuracy of CADD-based approaches by tapping into the continual supply of newly-deposited PDB entries to generate objective, large-scale tests for such approaches to increase their reproducibility and rigor, and by developing automated workflows to increase the scale and throughput of these validated CADD methodologies. The D3R’s Continuous Evaluation of Ligand Protein Predictions (CELPP) pilot initiative has already provided extremely encouraging results supporting the feasibility of the Center’s proposed strategy. Of particular note is the Center’s plan to develop a crowd-sourced ligand design game, analogous to the extremely successful and impactful

"FoldIt" crowd sourced protein design online game. This will, as noted, require and engender extensive and highly interactive engagement with the research community and potentially larger public in order to maximize both the necessary crowd-sourced input and the eventual impact of this approach. Moreover, the highly organized and selective use of social media such as email lists and Twitter is noted as a great strength. Another crucial aspect of the Center's emerging outreach is the effort to help establish data standards and methods standards for the user community, as is utilizing the resources and experience of Center members in data curation and management, and database development. Overall, resource dissemination is likely to continue be another great strength of the D3R.

Renewal Applications (if applicable):
N/A

CRITIQUE 2

Quality of the Community Engagement Plans:

The CE plan is divided into three aspects: blinded challenges, an annual workshop, and software hackathons.

Blinded challenges have a specific start time and provide participants (CADD method developers) with a way to test their ability to predict results from unpublished experiments. The program has already been running through a U01 mechanism and gets about 50 participants each year. This will be continued for on "approximately an annual basis for a least the first few years". It may be sunsetted based on feedback. If it were it would be in favor of a continuous challenge where the start date isn't set but participants can sign up on a rolling basis. Both of these approaches appear to be good ways for the community to get training and engaged. The plan would be better if more detail could be provided about the categories of participants that are expected to participate and the precise metrics that will be used to decide whether to move from one approach to the other. What will the evaluation metrics be?

An annual workshop will continue to be held – again it is presently funded from a U01. A clear description of who attends, how long the workshop is, who teaches at it are not provided. It is mentioned that the 2017 workshop was for 5 hours so presumably all were and will be 5 hours long? This seems like too short of a time for people from across the country to travel for. Again, there are no real metrics provided for how people will be recruited or the program evaluated. The virtual workshop seems to be popular according to a provided table with up to 126 participants, but did these folks stay on the phone the whole time or come and go? And there is little description about how to assess how well the participants learned or benefited from it.

Finally hackathons will be used. This is the least well-described component. For example, "*we plan to host, and have budgeted for at least intensive hackathons, held at least annually…*" There just isn't much description of this activity to fully understand the rationale, the audience, and how success will be measured.

Cost to participate is not described. Neither is recruitment. It does seem like the three activities are built to provide a range of training; however, there just isn't enough detail here.

Community Engagement Training:

Training will focus on "best practices for pose-prediction and affinity ranking using CADD methods". The training will be developed and led by D3R team members or community experts. It appears training will be done there computer modules that I assume are developed by team members? They will be piloted at the annual workshop. They will then go to the Biomedical Big Data Training Collaborative (BBDTC). Not much detail is provided about this other than it exists and was funded by a R25 grant. Electronic media will also be used to share information.

It would be most helpful to understand specific modules that will be developed, i.e., topics, how many per year, who is developing them exactly, etc. It would also be helpful to know how many users already exist for the BBDTC and what that demographic is. Altogether this component doesn't provide enough information to really understand what is going to happen. Clearly some training will be accomplished in the CE activities above as well.

Dissemination of Resources:

The center aims to disseminate both data and results through different mechanisms. The primary mode of results dissemination will be peer-reviewed articles. Note the DR3 challenges are already published in a special issue of Computer-Aided Drug Design and this will continue with funding of the center. This is a neat and innovative form of dissemination. A goal of annual publications or seminars delivered is not projected. Data will be disseminated in via a variety of web-based repositories. A website does already exist and shows a high number of page views since 2015. This is impressive.

Renewal Applications (if applicable):

ADDITIONAL COMMENTS:

The administrative team of Professors Rommie E. Amaro and Michael K. Gilson has already amply demonstrated that they have the expertise, capacity, and experience to effectively supervise and manage the D3R. Professor Burley's participation provides the needed direct administrative involvement of PDB. The qualifications of these three are simply outstanding. The assignation of three highly-qualified technical lead administrators (Altintas, Grethe and Hill) answering to the Co-Directors recognizes the need for a more focused supervision of the main driving technologies of the D3R, and the assignation of an experienced administrative assistant (Simas) is again sensible. The D3R also has the advantage of a strong central organization around UCSD campus' "Big Data freeway system". Some aspects of administration, such as the precise mechanisms for initiation or termination of DBPs or the pathways for potential conflict resolution, are not explicit in this proposal but are implicit in the architecture of the administration, e.g. in the annual EAC progress review or proposed ad hoc virtual meetings.

The EAC SAB membership represents the broad and comprehensive skillsets needed by the Center. The team of Challenge-oriented external evaluators to perform prediction evaluations and analysis is also a sensible augmentation to the advisory group, and will likely (continue to) provide valuable input in several areas of D3R development.

As discussed above, the resource has detailed and cogent sustainability plans.

DR3 will have two co-Directors – Professor Amaro and Gilson. Amaro is a rising star and has already served as the Director of the sunsetting P41, the National Biomedical Computation Resource. Gilson is an M.D./Ph.D. and leader in theoretical and computational molecular recognition. There is also a sub-contract that has its own PI – Ilkay Altintas. Several technical leads are listed. Given the experience of Co-Director Amaro with another P41 there is little doubt the group will have great experience for this proposed new center.

The plan would be improved by organizational chart that shows how the staff and components come together. How they will all interact is not described. No conflict resolution plan is provided and this really should be considered given the Co-Directing structure that is proposed. An operating procedure is described, but little to no governance detail is presented. Overall, there just isn't sufficient procedural or structural information presented here. Given that, it is a concern on whether the BTRR will function in an integrated and multi-disciplinary way. This is a sizeable weakness with the administrative plan.

A seven member EAC is described in the proposal and members are listed. They appear to come from across the US and from both industry and academia. The plan is for the EAC to meet once yearly and

individually as needed. During the annual meetings the EAC will review progress, provide guidance, and deliver feedback from community. The members selected for participation in the EAB seem appropriate for the center's mission. No discussion of how committee members will be rotated on an off is present. Overall the EAC is already in place. The application would be improved if it described how long the terms were and how new members will be selected. That said, it seems this detail is certainly manageable by the PIs.

The main plan for sustainability is that the core technologies will mature and that upon the maturation process will be less costly to maintain. It is also suggested that increased automation in future years will likely lead to cost reductions. Finally, as technology matures it will become shareable and other groups can run it on their own servers. That may be true, but there will still likely be the need for support of those users. Overall the sustainability plan could benefit from some further consideration.

THE FOLLOWING SECTIONS WERE PREPARED BY THE SCIENTIFIC REVIEW OFFICER TO SUMMARIZE THE OUTCOME OF DISCUSSIONS OF THE REVIEW COMMITTEE, OR REVIEWERS' WRITTEN CRITIQUES, ON THE FOLLOWING ISSUES:

COMMITTEE BUDGET RECOMMENDATIONS: The budget was recommended as requested.

Footnotes for 1 P41 GM135452-01; PI Name: Amaro, Rommie E

NIH has modified its policy regarding the receipt of resubmissions (amended applications). See Guide Notice NOT-OD-14-074 at http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-074.html. The impact/priority score is calculated after discussion of an application by averaging the overall scores (1-9) given by all voting reviewers on the committee and multiplying by 10. The criterion scores are submitted prior to the meeting by the individual reviewers assigned to an application, and are not discussed specifically at the review meeting or calculated into the overall impact score. Some applications also receive a percentile ranking. For details on the review process, see http://grants.nih.gov/grants/peer_review_process.htm#scoring.

MEETING ROSTER

Center for Scientific Review Special Emphasis Panel
CENTER FOR SCIENTIFIC REVIEW
PAR-17-316: Biomedical Technology Research Resource (P41)
ZRG1 BST-T (40)
06/19/2019 - 06/20/2019

Notice of NIH Policy to All Applicants: Meeting rosters are provided for information purposes only. Applicant investigators and institutional officials must not communicate directly with study section members about an application before or after the review. Failure to observe this policy will create a serious breach of integrity in the peer review process, and may lead to actions outlined in NOT-OD-14-073 at https://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-073.html and NOT-OD-15-106 at https://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-106.html, including removal of the application from immediate review.

CHAIRPERSON(S)

WYSOCKI, VICKI H, PHD
PROFESSOR
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY
OHIO STATE UNIVERSITY
COLUMBUS, OH 43210

MEMBERS

LARABELL, CAROLYN A, PHD
PROFESSOR AND DIRECTOR
NATIONAL CENTER FOR X-RAY TOMOGRAPHY LAWRENCE
BERKELEY NATIONAL LABORATORY
DEPARTMENT OF ANATOMY
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
SAN FRANCISCO, CA 94143

LEBRILLA, CARLITO B, PHD
DISTINGUISHED PROFESSOR
DEPARTMENTS OF CHEMISTRY AND BIOLOGICAL CHEMISTRY
UNIVERSITY OF CALIFORNIA, DAVIS
DAVIS, CA 95616

MUSEN, MARK A, MD, PHD
PROFESSOR
STANFORD CENTER FOR BIOMEDICAL INFORMATICS RESEARCH
STANFORD SCHOOL OF MEDICINE
STANFORD, CA 94305-5479

O'DONNELL, MATTHEW, PHD
PROFESSOR
DEPARTMENT OF BIOENGINEERING
UNIVERSITY OF WASHINGTON
SEATTLE, WA 98195

PLOEGH, HIDDE L, PHD
SENIOR INVESTIGATOR
PROGRAM IN CELLULAR AND MOLECULAR MEDICINE
BOSTON CHILDREN'S HOSPITAL
HARVARD MEDICAL SCHOOL
BOSTON, MA 02115

TROPSHA, ALEXANDER, PHD
PROFESSOR AND ASSOCIATE DEAN
ESHLEMAN SCHOOL OF PHARMACY
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
CHAPEL HILL, NC 27599

VAJDA, SANDOR, PHD
PROFESSOR
DEPARTMENT OF BIOMEDICAL ENGINEERING AND CHEMISTRY
BOSTON UNIVERSITY
BOSTON, MA 02215

WATKINS, SIMON C, PHD
DISTINGUISHED PROFESSOR, VICE-CHAIR, AND DIRECTOR
DEPARTMENT OF CELL BIOLOGY
CENTER FOR BIOLOGIC IMAGING
UNIVERSITY OF PITTSBURGH SCHOOL OF MEDICINE
PITTSBURGH, PA 15261

ZHENG, JIE J., PHD
PROFESSOR
STEIN EYE INSTITUTE
DEPARTMENT OF OPHTHALMOLOGY
DAVID GEFFEN SCHOOL OF MEDICINE AT UCLA
LOS ANGELES, CA 90095

MAIL REVIEWER(S)

BREY, WILLIAM W, PHD
ASSOCIATE SCHOLAR SCIENTIST
NATIONAL HIGH MANETIC FIELD LABORATORY
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FL 32310

CHAIT, BRIAN T, PHD
CAMILLE AND HENRY DREYFUS PROFESSOR
LABORATORY OF MASS SPECTROMETRY AND GASEOUS ION CHEMISTRY
ROCKEFELLER UNIVERSITY
NEW YORK, NY 10021

COON, JOSHUA J, PHD
PROFESSOR
DEPARTMENT OF CHEMISTRY
UNIVERSITY OF WISCONSIN, MADISON
MADISON, WI 53706

DUNN, KENNETH W, PHD
PROFESSOR
DEPARTMENT OF MEDICINE AND BIOCHEMISTRY
INDIANA UNIVERSITY SCHOOL OF MEDICINE
INDIANAPOLIS, IN 46202

FITZPATRICK, JAMES ALEXANDER, PHD
ASSOCIATE PROFESSOR
DEPARTMENT OF NEUROSCIENCE AND CELL
BIOLOGY AND PHYSIOLOGY
WASHINGTON UNIVERSITY SCHOOL OF MEDICINE
ST. LOUIS, MO 63130

GEE, JAMES C, PHD
ASSOCIATE PROFESSOR
DEPARTMENT OF RADIOLOGY
PERELMAN SCHOOL OF MEDICINE
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PA 19104

GRATTON, ENRICO, PHD
PROFESSOR AND PRINCIPAL INVESTIGATOR
DEPARTMENT OF BIOMEDICAL ENGINEERING
LABORATORY OF FLUORESCENCE DYNAMICS
UNIVERSITY OF CALIFORNIA, IRVINE
IRVINE, CA 92697

HAFENSTEIN, SUSAN, PHD
ASSOCIATE PROFESSOR AND DIRECTOR
CRYO-EM IMAGING FACILITY
DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR
BIOLOGY
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PA 16802

HEAD, MARTHA S., PHD
DIRECTOR
JOINT INSTITUTE FOR BIOLOGICAL SCIENCES
OAK RIDGE NATIONAL LABORATORY
OAK RIDGE, TN 37831

KENNEDY, MICHAEL A, PHD
PROFESSOR
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY
MIAMI UNIVERSITY
OXFORD, OH 45056

LAM, TUKIET T, PHD
DIRECTOR
MS AND PROTEOMICS RESOURCE
DEPARTMENT OF BIOPHYSICS AND BIOCHEMISTRY
YALE SCHOOL OF MEDICINE
NEW HAVEN, CT 06511

LIU, JUN, PHD
ASSOCIATE PROFESSOR
DEPARTMENT OF MICROBIAL PATHOGENESIS
MICROBIAL SCIENCES INSTITUTE
ADVANCED BIOSCIENCES CENTER
YALE UNIVERSITY
WEST HAVEN, CT 06516

LUBMAN, DAVID M, PHD
MAUD T LANE PROFESSOR
DEPARTMENT OF SURGERY
UNIVERSITY OF MICHIGAN MEDICAL CENTER
ANN ARBOR, MI 48109-0650

MARKO, MICHAEL, BS
ADJUNCT PROFESSOR
WADSWORTH CENTER
DEPARTMENT OF HEALTH NEW YORK STATE
ALBANY, NY 12201

MCSWEENEY, SEAN, PHD
DEPUTY DIRECTOR
PHOTON SCIENCES DIVISION
BROOKHAVEN NATIONAL LAB
UPTON, NY 11973

MECHREF, YEHIA, PHD
PROFESSOR AND CHAIR
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY
TEXAS TECH UNIVERSITY
LUBBOCK, TX 79409

MILLER, TIMOTHY A, PHD
ASSISTANT PROFESSOR
COMPUTATIONAL HEALTH INFORMATICS PROGRAM
BOSTON CHILDREN'S HOSPITAL
HARVARD SCHOOL OF MEDICINE
BOSTON, MA 02115

MORITZ, ROBERT L, PHD
PROFESSOR
PROTEOMICS RESEARCH
INSTITUTE FOR SYSTEMS BIOLOGY
SEATTLE, WA 98109

PANIN, VLADISLAV M, PHD
ASSOCIATE PROFESSOR
DEPARTMENT OF BIOCHEMISTRY AND BIOPHYSICS
TEXAS A&M UNIVERSITY
COLLEGE STATION, TX 77843

PAPADEMETRIS, XENOPHON, PHD
CONSULTANT
DEPARTMENTS OF RADIOLOGY, BIOMEDICAL IMAGING,
AND BIOMEDICAL ENGINEERING
YALE SCHOOL OF MEDICINE,
ENGINEERING, AND APPLIED SCIENCE
NEW HAVEN, CT 06520

PENG, JUNMIN, PHD
PROFESSOR
STRUCTURAL BIOLOGY AND DEVELOPMENTAL
NEUROBIOLOGY
SAINT JUDE CHILDREN'S RESEARCH HOSPITAL
MEMPHIS, TN 38105

PERRIMON, NORBERT, PHD
PROFESSOR
DEPARTMENT OF GENETICS
HARVARD MEDICAL SCHOOL
BOSTON, MA 02115

PETUKHOV, PAVEL A, PHD
PROFESSOR
DEPARTMENT OF MEDICINAL CHEMISTRY
AND PHARMACOGNOSY
COLLEGE OF PHARMACY
UNIVERSITY OF ILLINOIS, CHICAGO
CHICAGO, IL 60612

PUNSHON, TRACY, PHD
RESEARCH ASSISTANT PROFESSOR
DEPARTMENT OF BIOLOGICAL SCIENCES
DARTMOUTH COLLEGE
HANOVER, NH 03755

ROUT, MICHAEL P, PHD
PROFESSOR, HEAD OF LABORATORY
LABORATORY OF CELLULAR AND STRUCTURAL BIOLOGY
THE ROCKEFELLER UNIVERSITY
NEW YORK, NY 10021

SCHMID, MICHAEL F., PHD
SENIOR STAFF SCIENTIST
SLAC NATIONAL ACCELERATOR LABORATORY
MENLO PARK, CA 94025

SCHROEDER, WILLIAM J, PHD
PRESIDENT AND CEO
KITWARE, INC.
CLIFTON PARK, NY 12065

SHATKAY, HAGIT, PHD
PROFESSOR
DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES
UNIVERSITY OF DELAWARE
NEWARK, DE 19716

SHEKHTMAN, ALEXANDER, PHD
PROFESSOR
DEPARTMENT OF CHEMISTRY
UNIVERSITY AT ALBANY, SUNY
ALBANY, NY 12222

SMALHEISER, NEIL R, MD, PHD
PROFESSOR
DEPARTMENT OF PSYCHIATRY
COLLEGE OF MEDICINE
UNIVERSITY OF ILLINOIS AT CHICAGO
CHICAGO, IL 60612

STAGG, SCOTT M, PHD
ASSOCIATE PROFESSOR
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY
INSTITUTE OF MOLECULAR BIOPHYSICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FL 32306

STEMMER, PAUL M, PHD
ASSOCIATE PROFESSOR
INSTITUTE OF ENVIRONMENTAL HEALTH SCIENCES
WAYNE STATE UNIVERSITY
DETROIT, MI 48201

WANG, FENG, BS, PHD
PROFESSOR
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY
UNIVERSITY OF ARKANSAS
FAYETTEVILLE, AR 72701

WANG, JIN, PHD
ASSOCIATE PROFESSOR
DEPARTMENT OF CHEMISTRY
STATE UNIVERSITY OF NEW YORK AT STONY BROOK
STONY BROOK, NY 11794

WEMMER, DAVID E, PHD
PROFESSOR
DEPARTMENT OF CHEMISTRY
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CA 94720

YANG, XINMAI, PHD
ASSOCIATE PROFESSOR
DEPARTMENT OF MECHANICAL ENGINEERING
UNIVERSITY OF KANSAS
LAWRENCE, KS 66045-7618

YENNAWAR, NEELA H., PHD
DIRECTOR AND ASSOCIATE RESEARCH PROFESSOR
X-RAY CRYSTALLOGRAPHY FACILITY
AUTOMATED BIOLOGICAL CALORIMETRY FACILITY
HUCK INSTITUTES OF THE LIFE SCIENCES
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PA 16802

YU, HONG, PHD
PROFESSOR
QUANTITATIVE HEALTH SCIENCES
COLLEGE OF INFORMATION
AND COMPUTER SCIENCE
UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
WORCESTER, MA 01053

SCIENTIFIC REVIEW OFFICER

PETERSON, JOSEPH THOMAS, PHD
SCIENTIFIC REVIEW OFFICER
CENTER FOR SCIENTIFIC REVIEW
NATIONAL INSTITUTES OF HEALTH
BETHESDA, MD 20892

EXTRAMURAL SUPPORT ASSISTANT

RASULI, ZOHRA
DIVISIONAL LEAD ESA
CENTER FOR SCIENTIFIC REVIEW
NATIONAL INSTITUTE OF HEALTH
BETHESDA, MD 20892

Consultants are required to absent themselves from the room during the review of any application if their presence would constitute or appear to constitute a conflict of interest.