Office of Research and Development's Response to the Board of Scientific Counselors Report on ORD's National Center for Computational Toxicology (final report received September 2008)

February 2009

BOSC Computational Toxicology Subcommittee
Dr. George Daston (Chair)
Dr. James Clark
Dr. Richard DiGiulio
Dr. Muiz Mumtaz
Dr. John Quackenbush
Dr. Cynthia Stokes

Submitted by:
Dr. Robert Kavlock
Director, National Center for Computational Toxicology
Office of Research and Development

The following is a narrative response to the comments and recommendations of the BOSC review of ORD's National Center for Computational Toxicology (NCCT), held December 17 and 18, 2007, in Research Triangle Park, NC. The review was conducted by a standing subcommittee of the BOSC. The subcommittee had previously reviewed the NCCT in April 2005 and June 2006 and ORD responded to those reviews. In this third review, the BOSC noted, "…during the 2.5 years between its establishment and this review, NCCT has made substantial progress in establishing priorities and goals; making connections within and outside EPA to leverage the staff's considerable modeling expertise; expanding its capabilities in informatics; and making significant contributions to research and decision-making throughout the Agency." Furthermore they noted, "…many of the recommendations made by BOSC during its earlier reviews have been acted on by NCCT. This includes improved capabilities in bioinformatics through the funding of two external centers and in informatics and systems biology through staff hires; expansion of its technical approaches to even more programs within the Agency; and the formation of an extensive collaboration with the National Institute of Environmental Health Sciences (NIEHS) and the National Human Genome Research Institute (NHGRI) for its ToxCast^{TM} project."

Each charge question is shown below in bold, followed by the BOSC's comments in italics and ORD's response to the comments in regular type. A summary of the BOSC recommendations and ORD's responses is provided in Table 1 at the end of this report.

Charge Question 1: Does the scope and involvement of expertise in the project reflect activities consistent with the function of a Center?

The NCCT was founded only a few years ago and has been achieving a critical mass of expertise through selective hiring, external grants, and the formation of connections with other groups of experts within EPA. The purpose of this question was to gauge the progress of the Center in achieving the level of expertise needed to pursue its mission. The staff working in NCCT and those scientists involved from outside the Agency who are working as collaborators are highly qualified in various aspects of computational toxicology. The Center's effort to solidify formal agreements in terms of memoranda of understanding (MOUs), cooperative research and development agreements (CRADAs), etc., with various organizations has opened up a diversity of quality opportunities to leverage and enhance Office of Research and Development (ORD) efforts. A timely example is the February 14, 2008, announcement of the collaboration between NIEHS, NHGRI, and EPA's NCCT. As described in the press release, this collaboration leverages the strengths of each group to use high-speed, automated screening robots to test suspected toxicants using cells and isolated molecular targets.

The staff and collaborators at the center have the appropriate expertise and insights. The utility of the tools and deliverables can be enhanced if the staff moves toward being more explicit on how the tools under development support EPA risk assessments. Some of the ORD researchers seem to be searching for an application for their sophisticated tools, and discussions with Agency staff practicing risk assessments (Office of Pollution Prevention and Toxics [OPPT]; Office of Water, Office of Wastewater Management;

Office of Prevention, Pesticides, and Toxic Substances [OPPTS], etc.) could provide direction as to the appropriate milestones and deliverables for these efforts. The BOSC reviews and the Center would benefit if representatives from these Agency offices attended BOSC reviews to ensure that all parties understand how NCCT's efforts address the most relevant needs of the Agency. The BOSC wants to ensure that this advice is seen as encouragement to reach out to risk assessment practitioners. The ongoing work in developing the analytical approaches and information databases is of high technical quality, as the Center staff and collaborators are working on many new and exciting approaches. By holding research planning discussions with risk assessment practitioners, the applications of the computational toxicology tools and resources can be directed to ensure the most relevant and efficient use of data and models. (Recommendations #1 and #2 in Table 1)

ORD Response: ORD appreciates this recommendation. As noted in the report, NCCT regularly meets with program offices, risk assessors, and other potential practitioners in planning and conducting this research. A priority action item of the NCCT for FY2009 is to improve connectivity with NHEERL, NERL and NCEA relative to building the foundation for a transformation in the conduct of evaluating the toxicity of chemicals. We are continuing to engage Communities of Practice to help achieve this end. In previous reviews, some of these stakeholders were invited and attended meetings of this BOSC subcommittee. The next review will be a broad review of the computational toxicology program and the new implementation plan. For this and future meetings, Agency stakeholders will be invited to attend the meeting and enter discussions as appropriate. Further, the NCCT will ask such stakeholders to review and comment on the new implementation plan prior to the next BOSC meeting.

Charge question 1 continued:

One challenge for the center staff involved in developing informatics datasets will be to develop efficient and effective ways to handle the wealth of data available in some areas to avoid redundancies of data entries and to focus on the most informative data. Again, interactions with various program offices and their risk assessment activities should provide a basis to set the long-term goals for the Informatics/Data management team. This will allow the development of structured short-term and mid-term activities needed to meet the long-term goals. (Recommendation #3 in Table 1)

ORD Response: To address this important issue the NCCT has five main database-related, data-intensive projects: ACToR, ToxRefDB, ToxMiner, the ToxCast™ chemical registry, and DSSTox. ACToR (http://actor.epa.gov/actor) is the global repository of data that is relevant to environmental chemicals. It is populated from more than 200 public repositories of toxicity data to provide a broad, but in many cases shallow view of the universe of data available on chemicals of interest to the NCCT and the EPA. ToxRefDB is focused on extracting high quality *in vivo* toxicology data on chemicals in the ToxCast™ program, capturing study information down to the treatment group level, and extracting these into a relational database well-suited to predictive modeling. ToxRefDB is also being developed into a web-accessible resource that can be queried to derive

treatment related toxicity effects directly from the database. ToxMiner is a compilation of statistical tools capable of analyzing relationships between ToxCast™ and ToxRefDB data, and performing predictive signatures. The ToxCast™ chemical registry is used to track nominations for ToxCast™ screening, to track chemical procurement, sample identity and sample QC, and finally to link actual samples to ToxCast™ data. DSSTox is adding the quality reviewed chemical structure layer to data sets of interest to NCCT, and publishing additional inventories and toxicity data sets of interest to EPA and external groups. The underlying data model and database tables for all but DSSTox are being consolidated to remove data redundancy and to reduce the effort required to manage multiple related systems. DSSTox is primarily a file-based system, and as data are curated, they are entered into the ACToR system for further use. We are actively working with other partners (ORD, OPP, OPPT, OW, OHS, NCEA, the Tox21 partners) to prioritize chemicals to be entered into the system and to obtain and enter data. We believe this compilation of information on the toxicity of chemicals provides a solid foundation for the NCCT to not only understand the extent of public information on chemicals, but also to provide public access to this increasingly data rich repository of information on chemicals.

Charge question 1 continued:

The BOSC noted that it remains somewhat unclear how the Center intends to use ToxCast and associated analyses to approach risk assessment. For instance, species-to-species translation was mentioned, and the data are being obtained from multiple species, not just humans, but how the different species data will be reconciled was not discussed. Although the primary goal of the ToxCast project is prioritization of chemicals for detailed risk assessment, not the risk assessment itself, it is interesting to contemplate how the projected database and analysis might be directly relevant. Similarly, it was noted that an early decision regarding ToxCast was that ecology and paths of exposure were not going to be addressed in this project (at least not initially). Nonetheless, at several points, paths of exposure arose during the review because of their obvious relevance. The Subcommittee is prompted to ask how it might be addressed in future work. (Recommendation #4 in Table 1)

ORD Response: The NCCT has recognized the opportunity to address the full source-to-outcome continuum of risk assessment, and has recently done this in several ways. This need is reflected in the FY2009 priorities for NCCT that include increased connectivity with other components of ORD. Thus, NCCT has organized an ORD-wide workgroup to expand an overarching strategy for developing a high throughput approach to risk assessment-building from the example and lessons from ToxCast™ and expanding on applications to exposure and mode of action assessment. One part of this approach will be to develop exposure predictions on the thousands of chemicals relevant to ToxCast™, in a Center project tentatively called ExpoCast. Finally, the translation of ToxCast™ predictions directly to humans is being accomplished by direct comparison of results for rodent and human targets and pathways interrogated by complementary assays. In addition, a proposal has been accepted for consideration by the HESI Emerging Issues Program at its annual meeting in January 2009 to establish collaborations with the

pharmaceutical industry to supply chemicals with identified human toxicity for use in Phase IIb of ToxCast™. This phase would include at least 100 pharmaceutical compounds with known human toxicities and would extend ToxCast™ predictive signatures from Phase I of rodent toxicity endpoints, to similar toxicity endpoints in humans.

Charge question 1 continued:

The Subcommittee also noted that the means of using the eventual Virtual Liver models for actual risk assessment at EPA is unclear. The BOSC encourages additional thought and efforts along these lines, in collaboration with the appropriate EPA program office personnel. This is not a criticism of the current project vision by any means, but because direct or indirect application to risk assessment would be a fantastic result, it seems prudent to consider the possibility earlier rather than later. (Recommendation #5 in Table 1)

ORD Response: The Virtual Liver (v-Liver) is being developed in conjunction with NHEERL research activities. A detailed plan for v-Liver will be presented to the BOSC at the 2009 review. The objective of the v-Liver project is to coordinate an integrated *in vitro* and *in virtuo* program in the long-term for toxicity testing that is efficient, relevant to humans and less dependent on animals, with the ultimate goal of use in risk assessment. We agree that stakeholder involvement from EPA program offices is a critical requirement for the success of the v-Liver project. Although program office personnel were not directly involved in the early v-Liver research planning phase, senior scientists from NCEA/RTP, NHEERL and NCCT who have a good grasp of risk assessment needs for fulfilling EPA's mission, are part of the core team. Their collective insight into key challenges facing risk assessment and the requirement for future toxicity testing have been vital for shaping the vision for the v-Liver system. Therefore we believe the v-Liver project is poised to actively engage program office personnel to address challenges in mode of action (MOA) elucidation and quantitative dose-response prediction for chronic liver injury.

Program office personnel will be engaged in the design, development, and utilization of the system. This is being accomplished through a few practical use cases that demonstrate the value of Virtual Tissues for developing a proof of concept (PoC) for assessing the risk of environmental chemicals to liver physiology and human health. Over the next two years, the v-Liver PoC will define a subset of hepatic effects, apical endpoints, and relevant environmental chemicals which will be developed in close collaboration with program office personnel to ensure application to risk assessment and relevance to the EPA mission. In addition to developing a Virtual Tissues platform that will contribute in the long-term to the future of toxicity testing, the short-term milestones of the v-Liver PoC will also aim to address current client needs.

The v-Liver project plan (please see Appendix for outputs) outlines how stakeholders will be involved. Currently, the project is aligned closely with the ToxCast™, ToxRefDB and ToxMiner projects to develop methods to select environmental chemicals for the v-Liver

PoC focusing on nuclear receptor (NR) mediated hepatocarcinogenesis. Analyzing data from ToxCast™ and ToxRefDB has identified a range of pesticides and persistent toxic chemicals that match these criteria. Around ten chemicals will be used for the PoC and these will be selected in collaboration with program office personnel who are actively involved in their risk assessment and/or have substantial expertise in their MOA. We plan to develop these collaborations with stakeholders by providing them two main types of computational tools. In the short-term (FY09), interactive tools to aid hepatic MOA organization and analysis will be developed. In the medium (FY10) to long-term, these will be extended with prototype tissue-level simulation tools that will aid in investigating the quantitative relationships between MOA(s) and adverse effects.

The first deliverable for risk assessors is the v-Liver Knowledgebase (v-Liver-KB), which formally organizes information on normal hepatic functions and their perturbation by chemical stressors into pathophysiologic states. Information about hepatic physiology relevant for MOA analysis is dispersed across scores of public domain repositories as well as the biomedical literature and the v-Liver-KB will leverage semantic approaches, which are being increasingly adopted by the biomedical community, to provide effective tools that fill the gaps toxicity MOA organization and inference. The v-Liver-KB will be deployed as an interactive web-based and desktop tool to intuitively browse and query physiologic knowledge on PoC chemicals, to derive MOA(s) and to link assay results from ToxCast™, species-specific effects from ToxRefDB, and other evidence curated from the literature. We believe this system will provide computable information on key events that transparently indicate the uncertainties and data gaps and that make inferences on MOA from experimental data. In addition, we will work closely with risk assessors to customize the system for specific requirements. The v-Liver-KB will be deployed over the next two years and updated quarterly with any new information on the PoC chemicals.

The second deliverable (FY10), the v-Liver Simulator (v-Liver-Sim), dynamically simulates the key molecular and cellular perturbations leading to adverse effects in hepatic tissues. Initially, it will focus on modeling MOA leading to proliferative and neoplastic liver lesions at a hepatic lobular scale. The v-Liver-Sim is being developed as a cellular systems model of the hepatic lobule that will use MOA information from the v-Liver-KB to initially provide two outputs: the visualization of tissue changes at a histological scale and the assessment of lesion incidence. A version of this system will also be provided as a web-based/desktop tool to enable risk assessors to perform interactive and quantitative simulation of chemical induced perturbations of physiologic processes leading to toxic histopathologic effects. Eventually, the liver simulator will be integrated with PBPK models to model alternative exposure scenarios. Over the course of the project, the system will be evaluated in collaboration with risk assessors using PoC chemicals *in vitro* data from ToxCast™ and published *in vivo* data from rodents and humans.

Charge Question 2: Are the goals and milestones suitably described, ambitious, and innovative?

For the Center overall, the answer to this question is "yes." In particular, the goals of the Center are well-described, very ambitious and innovative, as well as important for the future of research at EPA. The issue of "milestones" is somewhat more complex, in part due to the varying levels of maturity for Center components. In most cases, previous accomplishments and current activities are well described, but more detail concerning projected future milestones would be helpful. It is recognized, however, that these projects are very innovative and substantial flexibility is appropriate. This is particularly true for less mature but highly creative projects such as the Virtual Liver and Virtual Embryo. Also, in considering goals and milestones, it may be appropriate to consider the timely integration of each project's accomplishments into the Agency's risk assessment activities. In the following paragraphs, Charge Question 2 is addressed in the context of the five major Center activities discussed at the review meeting.

ToxCast: Future plans for the project also are well described, although a more detailed time table for milestones past 2008 would be helpful. (Recommendation #6 in Table 1)

ORD Response: ORD agrees with this recommendation and has a more detailed timetable for ToxCast™ milestones which centers around the release of data, validation of predictive signatures, and generation of data as chemicals are tested in Phase II. With considerable data and experience now in hand from Phase I contractors and additional collaborations on the Phase I chemical library with laboratories within NHEERL and outside EPA, it will be possible to better articulate the directions for Phase II of ToxCast™. In addition, activities of the Tox21 consortium between NTP/NIEHS, NCGC/NHGRI and NCCT/ORD are maturing and beginning to identify near-term and medium-term goals. These activities will be described in greater detail in the second generation Implementation Plan, which we are now developing and will present to the BOSC at the next NCCT review. Please see appendix for detailed listing of milestones.

Charge question 2 continued:

IT/IM Activities: The project is highly and suitably ambitious, and its goals and substantial progress are well described. Again, future plans are described well in a general way, but more detail concerning future milestones (beyond 2008, which is well described) would be appropriate. (Recommendation #6 in Table 1)

ORD Response: Again, ORD agrees and has a detailed time table which emphasizes the deployment and continual upgrade of ACToR, integration of ToxCast™ and ToxRefDB *in-vivo* toxicology data, importation of available exposure, neurotoxicity, and reproductive toxicity data. A detailed listing of ACToR and ToxRefDB related milestones can be found in Appendix I.

Regarding ToxMiner, the first goal in FY09 is to incorporate all of the ToxCast™ Phase I data into ToxMiner. This involves processing the many individual data sets to eliminate

faulty data, to perform scaling and normalization, and to extract computationally useful parameters such as maximum effect levels and IC50 values. The second main task is to integrate the ToxMiner database with analysis tools for statistical analysis and machine learning. A third task is to integrate other biological information to help interpret the results of statistical analyses. In particular, we are incorporating pathway information and using this as an organizing principle to make sense of the results from the hundreds of individual ToxCast™ assays. The major goal of ToxCast™ Phase I is to develop a series of "signatures" linking *in vitro* data with *in vivo* toxicology. The related ToxMiner goal for FY09-FY10 is to produce and store these signatures and have them ready for validation on ToxCast™ Phase II chemicals. Planning is well underway for a ToxCast™ Data Summit in May 2009, which will provide a forum for external scientists to come and discuss alternatives for deriving predictive signatures of ToxCast™ HTS date relative to ToxRef identified phenotypes.

DSSTox will increase its interactions and alignment with major NCCT projects (ToxCast™, ToxRefDB, ACToR) and broader Agency and outside projects (NHEERL, OPPT, NTP, CEBS, EU REACH), providing key cheminformatics support, expanding DSSTox data file publications of toxicological data in support of predictive modeling, and enhancing linkages to resources such as PubChem for disseminating EPA, ToxRefDB and ToxCast™ bioassay results to the broader modeling community. Detailed milestones are found in Appendix I.

Charge question 2 continued:

Virtual Liver: Although narrower in scope than the foregoing projects, the Virtual Liver project is very ambitious; it also is relatively young, apparently becoming fully operational with the arrival of Dr. Imran Shah in September 2006. Its fit with the goals of NCCT is perhaps less clear than the previous two projects; it is more "visionary" in nature, and less directly applicable to risk assessment, as described by one of the EPA scientists involved. The goals of the project and the nature of research to be performed to achieve those goals are clearly described. There is some concern that this project may be overly ambitious. It may be helpful if key objectives were delineated and prioritized, perhaps indicating achievements that are critical to the success of the project and those that are highly desirable. Milestones for tracking the project's progress are not apparent, particularly in later years (3-5). This relatively young and very innovative project requires considerable flexibility, however, so the lack of detailed milestones in later years is very reasonable. (Recommendation #7 in Table 1)

ORD Response: The importance of developing and applying computational system level models of key phenotypic outcomes is reflected in the second goal of the new EPA Strategic Plan for Evaluating the Toxicity of Chemicals that is currently working its way through final concurrence by the Agency. NCCT recognizes the need to better delineate the goals and milestones of the v-Liver project, and we have made this a key activity in response to the comments of the BOSC. NCCT is convinced the future of toxicology will be heavily dependent upon the development of computational systems level models and has played a key role in the development of this plan and its execution through this

project. Current and additional details will be provided at the next review of the BOSC. The short-term goals for the v-Liver project are to identify environmental chemicals for the PoC system. Once there is buy-in from EPA stakeholders (program offices and NCEA) on these chemicals, the team will begin populating the v-Liver-KB with relevant mechanistic and MOA information on these chemicals including *in vitro* data from ToxCast™ and *in vivo* data from the literature. Concurrently, the team will develop a prototype virtual hepatic lobule to understand the key cellular responses necessary for modeling cancer progression beginning with nuclear receptor activation. Data generated by ToxCast™ as well as external collaborators/new contracts will be used to begin quantitative parameterization of the cellular and molecular responses, and their evaluation using published *in vivo* rodent data. The detailed milestones for the project are described in Appendix I.

Charge question 2 continued:

Developmental Systems Biology (Virtual Embryo). This project is at a substantially earlier stage than the Virtual Liver project; it is led by Dr. Thomas Knudsen who joined NCCT in September 2007. The issues of goals and milestones are essentially the same as for the Virtual Liver, that is, strong on the former, but understandably weaker on the latter. It is the Subcommittee's expectation that a more concrete research plan with goals and milestones will be developed over the coming months. (Recommendation #8 in Table 1)

ORD Response: A formal research plan for the Virtual Embryo, including goals and milestones, has been developed. The long-term goal will provide a computational framework that enables predictive modeling of prenatal developmental toxicity. The project is motivated by scientific and regulatory needs to understand how chemicals affect biological pathways in developing tissues, and through this knowledge a more ambitious undertaking to predict developmental toxicity. The research plan is built on an expanded outlook of experimental-based techniques that aim to identify 'developmental toxicity pathways' and an expanded scope of computational search-based techniques that apply such knowledge into models for chemical dysmorphogenesis. Dr. Knudsen, the lead scientist for this program, was recently invited to NCEA where he provided an overview of the project. This has led to close coordination between the computational models and the risk assessment priorities.

Virtual Embryo's short-term goals address the knowledgebase (VT-KB) and simulation engine (VT-SE) to enable *in silico* reconstruction of key developmental landmarks that are sensitive to environmental chemicals. Initial research focuses on early eye development. Proof-of-principle (2yrs) will be measured by high fidelity simulation models to demonstrate several generalized principles, including the ability to reconstruct genetic defects *in silico*, classify abnormal developmental trajectories from genetic network inference, and predict teratogen-induced defects from pathway-level data. A much more detailed research plan will be provide to the BOSC in its 2009 review of the NCCT, and detailed examples of current envisioned milestones are found in Appendix I.

Charge question 2 continued:

Arsenic BBDR: This project is unusual among NCCT projects in that it is oriented toward a specific chemical with a specific issue (Safe Drinking Water Act revisions) rather than an approach developed with diverse chemicals in mind. However, this project is likely to inform the eventual development of other biologically-based dose-response models and their application to risk assessments by the Agency. Thus, in addition to informing the controversial issue of arsenic risk assessment, the project is more broadly relevant to the mission of the NCCT. The goals of the project are very clear and well described. Milestones, however, are not stated, and may be particularly important for this project, which has a clear deadline (2011) in order to be useful for the 2012 Safe Drinking Water Act review cycle. (Recommendation #9 in Table 1)

ORD Response: At the time of the BOSC review in December, 2007, considerable effort had been devoted to planning the development of a BBDR model for carcinogenic effects of inorganic arsenic (iAs). The initial focus of the planning process was a literature review to identify data needs. This review had shown that the pharmacokinetics (PK) of iAs were relatively well-studied, though there were some significant remaining PK uncertainties. The literature was not, however, sufficient to identify with any confidence the relevant mode or modes of action (MoA) of iAs responsible for its carcinogenic effects. We therefore developed a generic experimental design that focused on: (1) the description of a potential MoA as a sequence of key events; and (2) experimental characterization of the dose-time response surfaces for the key events. For any given candidate MoA, it was anticipated that this experimental approach would have provided sufficient data to allow ranking of candidate MoAs by dose and time course. The MoA or MoAs acting at the lowest doses and earliest time points would be considered to be the drivers for the apical cancer outcomes.

The next step in the process was to elicit research proposals from NHEERL iAs researchers that were to be based on the suggested experimental approach for characterizing candidate MoAs. The literature is consistent with a relatively large number of MoAs for iAs. These include (among others) oxidative stress, cytolethality and regenerative cellular proliferation, altered patterns of DNA methylation, altered DNA repair, and DNA damage. Receipt of the proposals was followed by an external peer review meeting. The outside experts judged that the proposals received did not adequately represent plausible modes of action, which caused NHEERL management to markedly reduce the planned BBDR modeling effort and focus on-going research on iAs PK, with particular emphasis on evaluation of the arsenic 3-methyl transferase knockout mouse. The NCCT involvement in the arsenic mode of action BBDR models has been redirected to stronger interactions with existing NCCT projects in ToxCast^{TM} and the v-Liver, and will be presented to the BOSC at its next review of the Center.

Charge Question 3: Are there significant gaps in the approach that can be pointed out at this point in the evolution of the project?

ToxCast: Specifically, the Subcommittee notes that the structural specification of the database for compilation and rigorous quantitative analysis of the ToxCast data remains unclear. Because the data types are highly heterogeneous and the dataset is very large, developing these structural specifications will be a challenge that the Subcommittee suggests should be addressed as soon as possible. The IT/IM team acknowledges that this area is a significant challenge (e.g., the description in the write-up provided to the Subcommittee prior to the review meeting). One suggestion is that the ToxCast team compiles a list of some specific use cases, for example, specific questions that they intend to address with the database. This will help make concrete the needed database attributes that will allow the analysis for the chemical prioritization that is the end goal of the ToxCast project 1). (Recommendation #10 in Table 1)

ORD Response: Over the last several months, these issues have become clearer, mainly due to the fact that we now have access to large parts of the ToxCast™ data. With the exception of the microarray genomics data, which has been delayed due to lack of consensus on the most appropriate bioasssy conditions, the results of all of the assays can be reduced to a small number of summary parameters. In most cases, one of these will be a characteristic concentration for each chemical in the assay (EC50, IC50, lowest observed concentration at which a significant effect is seen). The second parameter will often be a magnitude of response. For all of the assays, we can extract a relevant concentration and for many, a response magnitude. Related to this, the endpoint data we will be predicting from ToxRefDB are characteristic concentrations, which are the lowest doses at which a particular effect was seen with statistical significance. A third variable in some assays is time – cell based assay data in some cases is provided at 2-3 time points (e.g. 6, 24 and 48 hours). We track these times, but treat each of the times as separate assays. Finally, most assays can be linked to biological pathways, either directly through the gene or protein, or through a higher-order processing being probed. Although ToxCast™ was envisioned to support chemical prioritization efforts of Agency regulator offices, it has since been viewed as a source of ancillary information that can be used in evaluating risks. Examples of this include interest of the toxic substances office on the effects of perfluoroacids, NCEA with phthalates, and the pesticide office with conazoles. Such interest demonstrates the multiple values the information emerging from ToxCast™ is having on the regulatory programs of EPA beyond chemical prioritization. We anticipate continued interest in the use of ToxCast™ in risk assessment considerations and are engaging NCEA in optimal ways to bridge the applications.

As already stated, the goal of ToxCast™ Phase I, as supported by the ToxMiner system, is to find links between *in vitro* assays and *in vivo* toxicity as captured in ToxRefDB. These can be statistical correlations or more biologically-based toxicity pathway linkages. Given this, the ToxMiner database has been organized into five main pieces:

1. Chemical information – this holds chemical identity and structure
2. Assay information – this holds the summary values extracted from in vitro assays and from ToxRefDB (concentrations, response magnitude), as well as other related quantitative and qualitative information on chemicals such as physico-chemical properties and chemical class information.
3. Data preparation – for many of the data sets, several pre-processing steps need to be undertaken to map raw data onto the canonical chemical and assay data structure. These tables and data structures enable these steps to be carried out in well-controlled manner
4. Statistical analysis workflow – many calculations need to be carried out to find signatures and the results need to be tracked and made available to the ToxCast™ team on the web. We are implementing specific data tables and code to carry out these steps.
5. Pathway information – this set of data tables and tools are being designed to allow the analysis of the ToxCast™ data in terms of biological pathways.

Charge Question 3 continued:

IT/IM Activities: The major gap noted for this activity was described in the ToxCast project section above. In addition, finding an efficient and effective methodology for extracting data from text sources was a concern for the Subcommittee. A trial of natural language processing (NLP) for pulling information into some of the databases was described. The Subcommittee notes that this method has been attempted rather unsuccessfully by various research groups over probably 2 decades and thereby encourages the exploration of other possible approaches as well. (Recommendation #11 in Table 1)

ORD Response: NCCT agrees and is developing two main uses for literature mining, for which we believe current technology is suitable. In the first case, we need to extract tabular data for use in ToxCast™ and the virtual tissue project. These are, for instance, quantitative values associated with *in vivo* toxicity or *in vitro* assays. Here we are using text mining as a sophisticated version of a PubMed search to prioritize documents for data extraction and to do an initial automated data extract. The results are then presented to an analyst to do manual quality control and data cleaning.

The second task is to generate hypotheses about biological processes such as the co-occurrence of gene expression changes and the observation of higher-order phenotypes. The lack of success that the reviewer alludes to, we would argue, is in taking these hypotheses and assigning some truth value to them based on statistical arguments. We are using these simply as starting points for building representations of pathways and processes that will be tested through further experiments and analyses. A more detailed explanation of our approach to literature mining and evidence of utility will be presented at the next BOSC review.

Charge Question 3 continued:

Virtual Liver: Dr. Shah and his group are commended for having a good command of the significant breadth of biology, toxicology, and modeling that impacts the project. In addition, the "big picture" vision described is useful—there are many important questions in the field and not limiting the vision too early is appropriate. The Subcommittee believes that this should be balanced, however, with some very specific goals, milestones, and timelines for the next few years that are clearly attainable with the resources at hand in order to assure some useful concrete outcomes. In a project with this possible magnitude, it can be tempting to try to do everything, both in terms of the various project approaches (knowledgebase (KB), biological modeling, dosimetry modeling, etc.) as well as the scope within any one approach (breadth of the KB, breadth and detail of every model, etc.), and thereby end up with little actually completed. One suggestion is that Dr. Shah and the group develop a short prioritized list of specific scientific research questions relevant to EPA's goals that they desire to address as soon as possible, and use this to focus first iterations of development of both the KB and model(s). More explicit milestones and goals for these highest priority questions then can be developed. Later iterations of KB development and modeling can add scope (breadth/depth) to allow NCCT to address additional research questions. (Recommendations #6 and #12 in Table 1)

ORD Response: The question, "*How can in vivo tissue level adverse outcomes in humans be predicted using in vitro data?"* is the "Grand Challenge" scientific problem in toxicology that motivates the v-Liver project. This is a very ambitious goal and infeasible to achieve in the broad sense in just a few years. Hence, the v-Liver project will take a few steps towards realizing this long-term objective by focusing on a tractable proof of concept (PoC) system using ten environmental chemicals that activate nuclear receptors and cause a range of apical effects in cancer progression (non-proliferative lesions, pre-neoplastic lesions, and neoplastic lesions). The project will engage program office personnel to ensure relevance to EPA's mission and provide deliverables for risk assessment within the first two years. These deliverables focus on two main scientific questions:

a) How can tissue level adverse effects be modeled to enable extrapolation? The v-Liver leverages the Mode of Action Framework and public sources of mechanistic information to formalize the description of key events leading to adverse hepatic outcomes. Our claim is that MOA knowledge can be universally described across species, organs, chemicals and doses, using genes, their interactions, pathways and cellular responses that lead to toxic effects. This claim will be tested in the PoC by: (a) organizing sufficient information about the 20 nuclear receptor-activators to demonstrate that key events in the MOA(s) can be described generally for extrapolation across chemicals and species, and (b) using semantic methods to build an ontology for the physiologic processes, a knowledgebase to integrate this information, and inference tools for extrapolation. The result of this exercise will be delivered as the v-Liver-KB.

b) How can the tissue level effects be extrapolated across doses and time? Our claim is that quantitative tissue level effects can be generated from qualitative logical descriptions of the MOA(s), chemical-specific data for key events and simulation of the tissues as a cellular system. The rationale for the v-Liver Simulator is to implement a virtual hepatic lobule as a complex cellular system to investigate emergent tissue-level effect due to alternative MOA(s) at very low environmentally relevant doses. To extrapolate between species, chemicals and doses, the v-Liver team is collaborating across ORD and extramural funding to develop *in vitro* models and assays to relevant quantitative data key events. In addition, to estimate internal dose and to model alternative exposure scenarios the project is working closely with PBPK modeling efforts across ORD. The deliverable for this part of the project will be the v-Liver Simulator.

Charge Question 3 continued:

Virtual Liver: The Virtual Liver activity will result in models of parts of the biology being developed simultaneously and presumably by different individuals. Because the idea is to integrate these models eventually to predict effects from molecular function to physiologic outcome, the compatibility of the models is paramount. Dr. Shah indicated that he is cognizant of and planning to manage this issue, for instance, by looking into the efforts of the international Physiome Project. The Subcommittee members note that, to their knowledge, the issue of common coding language, which has been addressed quite extensively by the Physiome Project, does not appear to have addressed more subtle but critical compatibility issues concerning biological and mathematical specifications among models, such as compatibility of assumptions, equilibrium approximations, time scales, and so forth. Hence, beyond managing compatible coding, the activity group is encouraged to actively plan for and manage on an ongoing basis the specifications that must be shared among models so as to produce compatibility when it is needed. (Recommendation #13 in Table 1)

ORD Response: This is indeed a difficult and very important issue to consider. To this end NCCT is beginning to address the issue on two fronts:

1. NCCT plans to raise this issue for discussion by multi-scale modeling experts at the NCCT organized International Workshop on Virtual Tissues, to be held in Research Triangle Park, NC, April 21-23, 2009. This workshop will have representation from the Physiome project and the SBML project and is co-sponsored by the European Union.
2. In addition, the NCCT is actively collaborating with PBPK modelers in the Agency to develop a formal specification that will ease the integration with v-Liver-Sim. The effort is using semantic technology to define physiologic models at the organism level that can interface with existing tools in NERL.

These two integrated efforts will be important early steps for addressing this problem.

Charge Question 3 continued:

Virtual Embryo: Because the data needs of the proposed models may be significant, the Subcommittee notes that it will be critical to identify and enlist appropriate supporters and collaborators to provide such data. The track record of the principal investigator suggests that this will develop naturally. (Recommendation #14 in Table 1)

ORD Response: With successful proof-of-principle (2 yrs), the computational model of early eye development will be used to create general models of morphogenesis during subsequent years. Any proposed model of chemical dysmorphogenesis must be sufficiently abstract to be computationally feasible and yet detailed enough to enable the realistic expression of developmental defects across chemicals, doses, tissues, stages, and species. The data needs of the proposed models will be significant as noted by the Subcommittee. Preliminary computational models can attach existing data from *in vitro* studies and semi-arbitrary parameters from *in silico* resources. These models will be calibrated across species (zebrafish, mouse, rat, human) and tested for predictive capacity. In this regard, the Virtual Embryo will leverage data generated by NCCT's high-throughput chemical screening and prioritization research program (ToxCast™, ToxRefDB) to model developmental toxicity pathways.

Importantly, to stimulate research in this area, NCER released a funding opportunity under its Science To Achieve Results (STAR) research program, "Computational Toxicology Research Centers: *in vitro* and *in silico* models of developmental toxicity pathways" (EPA-G2008-STAR-W). Collaboration with future STAR center(s) can provide experimental data to identify developmental toxicity pathways and computational models for developmental defects.

Because conservation of cell signaling is a founding principle of early development across species and stages, the *in silico* toolbox is likely to be extensible across morphoregulatory responses. As such, *in silico* models built from scratch can be generalized to other systems (neural tube, cardiac, urogenital) and alternative models (embryonic stem cell assays, zebrafish embryos) for chemical-pathway interactions. In this regard, the Virtual Embryo has begun to identify and enlist collaborators at NHEERL to help provide such data.

High-throughput platforms now offer a powerful means of data gathering to discover key biological pathways leading to apical endpoints of toxicity, and computational model structures our ability to integrate these data across biological scales to build predictive models that address mode-of-action. Successful computational models can become increasingly important in EPA efforts to translate pathway-level data into risk assessments, and in that regard the Virtual Embryo has also begun to identify and enlist support from NCEA. A web-site has been developed to communicate publically about the project (http://www.epa.gov/ncct/v-Embryo/).

Charge Question 3 continued:

Arsenic BBDR: The Subcommittee encourages continuous communication with the appropriate program office personnel so that concerns, objections, and skepticism can be addressed early and on an ongoing basis. The group is commended for having such communication already in place and it is encouraged to maintain that communication to the greatest degree possible. (Recommendation #15 in Table 1)

ORD Response: As discussed in the response to charge question 2, this project was largely terminated in 2008, with the exception of a few smaller efforts on pharmacokinetics of arsenic. NCCT efforts are being redirected to incorporate concepts of BBDR in the virtual tissue models, particularly from the viewpoint of dose-response extrapolation. Additional NCCT efforts are being directed at interpreting the results of ToxCast™ *in vitro* concentration responses relative to the range of potential external exposures that could provide equivalent tissue level responses (i.e., reverse toxicokinetics). As we move forward in these areas, we will ensure adequate discussion with client offices in EPA takes place on a routine basis.

Charge Question 4: Does the work offer to significantly improve environmental health impacts and is the path toward regulatory acceptance and utilization apparent?

ORD Response: ORD is very appreciative of the committee's affirmation of work and progress in ToxCast™, Informatics, and the virtual tissues. The NCCT will present further updates on progress at the next committee review.

Charge Question 4a: In addition, specifically for the Arsenic BBDR project: Does the proposed computational model have the potential to identify and reduce uncertainties with the risk assessment process?

The answer to this question is yes, depending on data gaps identified and resources made available. This study might not give all the answers but will get us halfway there. EPA recognizes that developing a universal arsenic model describing several cancer endpoints is a formidable challenge. Hence a step-wise research project with an eye for the future is proposed. Initially, a generic model for cancer will be developed that will incorporate key steps of the mode of action commonly shared for multiple cancer types such as oxidative stress. This model, in turn, will serve as an engine to develop specific cancer models as the need arises and resources become available. To ascertain whether appropriate steps are being incorporated, a thorough literature review of experimental and epidemiological data and expert consultation has been proposed. It also is acknowledged that even though there is a lot of data, they are somewhat weak to generate exposure time course response curves. Appropriate experiments have been proposed to fill the research needs to develop a realistic model.

ORD Response: Please see earlier response regarding the arsenic BBDR project.

Charge Question 4b: Will the model be able to help identify susceptible populations and compare potential risks in those populations with less susceptible populations?

Yes, the initial generic model development exercise will allow identification of issues such as mechanisms that operate in general versus subpopulations, such as susceptible populations with varying degree of arsenic methylation. Such issues could be the subject of workshops to explore the issue of the extent of polymorphism in the human population.

The short-term (1-2 years) goal is the establishment of a coordinated program of laboratory research to generate essential data needed to develop a BBDR model that will increase confidence in the predictions. To start with, the model development will be initiated with available data. Work proposed includes multistage clonal growth modeling, target tissue dosimetry, and methylated metabolites of arsenic.
The long-term (3-5 years) goal of developing a robust version might be too optimistic. As the project gets underway, new questions and issues might be identified that will require additional laboratory research and continued resources. The project has a good future as it can be easily adapted to the latest (2007) National Academies toxicity testing report that recommends a systems biology and computational tool integration.

ORD Response: Please see earlier response regarding the arsenic BBDR project.

Charge Question 4c: Is coordination between model development and associated data collection sufficient to avoid problems with models being either over- or under-determined?

Yes, it is desirable to see what health effects are caused at lower doses to avoid the potential of compromise in setting an arsenic standard based on cost-benefit analysis.

ORD Response: Please see earlier response regarding the arsenic BBDR project.

Charge Question 5: Have appropriate data management and analysis tools been incorporated into the project?

ToxCast™: The construction of the warehouse remains an open question. Ultimately, a database is a model of the interactions that exist in the underlying data and the relationships relevant to the analysis that will be performed. The diversity of the data, representing a wide range of in vivo and in vitro assays from multiple species, makes building such a model a significant challenge. The project seems to be lacking a set of analytical objectives necessary for building the relevant use cases that ultimately will inform the process of database construction, and this ultimately will determine its utility. At this stage, ToxCast needs to begin to define analytical outcomes in order to set goals and milestones with regard to developing and validating analytical protocols. This is an essential step at this point as it will help to anchor future development and make it relevant. This also will help to define the requirements of the interfaces that are built to access the data.

Further, the ToxCast group should be encouraged to release the data and databases at the earliest possible time and to consider a "CAMDA-like" workshop in which the research community is offered access to the data with the challenge of using the data to effectively predict end points. At least three advantages to the program will be derived from these efforts. First, public release will help to drive the creation of relevant use cases that will further database development. Second, it will assist in evaluating data access protocols and tools to assure the greatest utility to the research and regulatory community. Third, it will accelerate the development of predictive algorithms to combine the data to make predictions about relevant phenotypic outcomes. (Recommendations #6, 10 and 16 in Table 1)

ORD Response: The first part of this question (database design and construction) was addressed in the response to charge question 3. The ToxMiner database is able to capture and provide all of the summary information which we believe is going to be useful for statistical and pathway-based analysis of the ToxCast data sets.

The second question relates to analytical outcomes. By this we assume the reviewer means the desired outcomes of analyses of the ToxCast data. We believe that the outcome of ToxCast will be a series of well-defined procedures that take as input the results of a set of in vitro assays run on a chemical and give a result which is a statement about the likelihood that the chemical will lead to a particular toxicity phenotype. The simplest procedure is a formula (e.g. a logistic regression model) that uses the IC50 values for several assays and gives a binary prediction for a particular toxicity. More complex procedures would use the results from a set of assays to predict whether a particular pathway is activated. Then we could have a function that predicts the likelihood of the outcome, given the activation of one or more pathways. The current database has been designed to hold both the numerical data required to test these models, and the model parameters and outcomes. In summary, we feel that this issue has in general been resolved over the last several months although many details still need to be worked out, particularly regarding the best statistical approaches to be used, and the precise way that pathway information will be incorporated.

With regard to the last comment by the reviewer, a recommendation that we hold a CAMDA-like workshop, we are currently planning such a meeting to be held in May 2009. We plan to make all of the ToxCast™ data available to analysis partners in early 2009. By having a larger community trying many analysis techniques on this data, we will maximize our chances of success.

Charge Question 5 continued:

V-liver: With regards to populating the KB, the use of NLP probably is not the best solution. NLP does not work well with the scientific literature, and its application in this domain remains an area of active research. Application of NLP has the potential to introduce a great deal of noise in the system, leading to many potential false associations that could lead to more problems than it solves. Consequently, other methods, including expert or community curation, should be explored.

On a larger scale, the greatest potential problem will be linking each of the domain-

specific models to build a predictive system. Again, this remains an area of active research and one that may present significant barriers to developing verifiable solutions. The greatest challenges will be to validate any models that emerge from the analysis. Finally, there is a need to develop standards for interactivity and try to interface with developing standards within the community. (Recommendation #17 in Table 1)

ORD Response: Linguistic resources have several applications in the Virtual Embryo although an important challenge noted by the Subcommittee is to unambiguously code unstructured (text) data in a form that can be processed by a computer to derive interesting relationships and causality. Querying within the proper context can make these more precise and less noisy. NLP enhances the coarse semantic search for specific concepts and then provides a way to automatically extract the key facts, relationships and quantitative information. The results are then presented to an analyst to do manual quality control and data cleaning. As such NLP extends, but does not replace the need for a formal concept model (ontology) to organize the relevant information about developmental processes and toxicities that is often present in literature in an unstructured format.

Also noted by the Subcommittee, a broader network of expertise within the developmental toxicology community may be useful to building the information network. Virtual Embryo has incorporated two open ontologies to arrange information, one for embryology and the other developmental toxicology, and implemented this ontology in Protégé (http://obofoundry.org/). This formal ontology will be available for community participation in linking each of the domain-specific models to build a predictive system for the embryo as a whole. Furthermore, informal ontologies that include less explicit information about a pattern of malformations and underlying embryology can make a useful contribution when the end-user is knowledgeable about the field. Hence, Virtual Embryo is piloting a Wiki-space (http://v-embryo.wikispaces.com/) to generate hypotheses about the co-occurrence of specific malformations to common embryology, or the relationship of genetic defects to higher-order phenotypes, for building representations of pathways and processes that can be tested through further experiments and analyses.

Charge Question 5 continued:
V-Embryo: It remains to be seen how well it will eventually integrate with the overall program, and its integration with other internal and external initiatives needs to be resolved. Nevertheless, it appears that this project could provide an opportunity to explore the results emerging from ToxCast, and it may help direct selection of the next generation of compounds for analysis in ToxCast. (Recommendation #18 in Table 1)

ORD Response: Although still early in its development Virtual Embryo has begun to integrate with other activities, especially ToxCast™ and the Virtual Liver. Since its inception last December and the review addressed here, the v-Embryo has been:

1. integrated into NCCT's Computational Toxicology Research second generation Implementation Plan;

2. presented at five seminars at EPA (including NCEA) and six seminars outside EPA (including a Gordon Research Conference);
3. introduced at NCCT's Computational Toxicology education course, at two presentations describing the implementation of prenatal developmental studies in ToxRefDB (manuscript in preparation), and one presentation on ToxCast^{TM}'s NovaScreen assay (manuscript in preparation);
4. the topic of one book chapter (in print) and seven abstracts (five in print and two accepted);
5. reflected in one submitted abstract in collaboration with Virtual Liver, and three
6. submitted abstracts in collaboration with ToxCast^{TM}; and
7. presented in the virtual tissue research section at the Human Health Program Review (BOSC, January 2009).

.

Appendix I: Summary Action Items

Detailed Milestones in response to Charge Question 2

ToxCast™:

FY09

- First initial publications and public access to ToxCast™ *in vitro* assay data
- Completion of generating all of the ToxCast™ Phase I data
- Sharing of ToxCast™ Phase I data with data analysis partners and hosting of the first "ToxCast™ Data Analysis Summit"
- Develop a series of "signatures" linking ToxCast™ *in vitro* data with ToxRefDB *in vivo* toxicology.
- Initiate generation of ToxCast™ Phase II data
- Quarterly public releases of new ToxCast™ data of various study types

FY10

- Quarterly public releases with new ToxCast™ data
- Completion of generating all of the ToxCast™ Phase II data
- Sharing of ToxCast™ Phase II data with data analysis partners and hosting of the second "ToxCast™ Data Analysis Summit"
- Validation of predictive "signatures" linking ToxCast™ *in vitro* data with ToxRefDB *in vivo* toxicology

FY11

- Quarterly public releases with new ToxCast™ data
- Application of toxicity predictions from Phases I and II of ToxCast™ to chemical prioritizations in EPA Program Offices
- Initiate generation of ToxCast™ Phase III data on chemicals and nanomaterials requiring prioritization

ACToR:

FY09

- Initial public deployment
- Significant version 2, including refined chemical structure information
- Develop workflow for tabularization of data buried in text reports
- Integrate all ToxCast™ and ToxRefDB data
- Quarterly releases with new ToxCast™ data

FY10

- Quarterly releases with new ToxCast™ data
- Implementation of a process to gather tabular data on priority chemicals from text reports
- Perform survey of sources of exposure data and import any remaining sources
- Develop flexible query interface and data download process
- Develop process to extract data from open literature

FY11

- Quarterly releases with new ToxCast™ data

ORD Response to BOSC Computational Toxicology Letter Report
February 2009

ToxRefDB:

FY09

- Initial public deployment of chronic toxicity data
- Public deployment of reproductive and developmental toxicity data
- Develop flexible query interface and data download process
- Develop workflow for curation of similar, but non-guideline chronic, reproductive and developmental study types
- Public deployment of developmental neurotoxicity data
- Quarterly public releases of new data of various study types

FY10

- Quarterly releases with new ToxCast™ data
- Implementation of a process to curate data on ToxCast™ Phase II chemicals from multiple sources

FY11

- Quarterly releases with new ToxCast™ data

DSSTox:

FY09

- Publish paper and property files on ToxCast 320 chemical inventory, with guidance for SAR modeling study
- Publish DSSTox ToxCast 320 categories file and DSSTox ToxRef summary data files
- Coordinate efforts to structure-annotate and provide effective linkages to microarray data for toxicogenomics
- Compile and publish public genetic toxicity data and SAR predictions for ToxCast 320
- Restart Chemoinformatics Communities of Practice using EPA's Science Portal;

FY10

- Publish new DSSTox database and doc
- Explore new approaches to SAR modeling based on feature categories within existing DSSTox files and ToxCast™ data
- Expand CEBS collaboration to incorporate DSSTox chemical content, create chemical linkages to external projects;
- Separately publish DSSTox structure inventory with various chemical classifications for use in modeling using publicly available tools

FY11

- In collaboration with ACToR, establish procedures and protocols for automating chemical annotation of new experimental data submitted to CEBS or NHEERL
- Document and employ PubChem analysis tools in relation to published DSSTox and ToxCast™ data inventory in PubChem
- Collaborate with SAR modeling efforts to predict ToxCast™ endpoints using in vitro data

- Continue expansion of DSSTox public toxicity database inventory for use in modeling with co-publication and linkage to ACToR and PubChem

v-Liver:

FY09

- *Prioritize proof of concept (PoC) environmental chemicals with clients.* Using toxicity data from ToxRefDB and bioactivity data from ToxCast^{TM}, a subset of Phase I chemicals will be selected for the PoC, which will be finalized in collaboration with program offices to ensure relevance to EPA needs.
- Begin deployment of v-Liver KB on physiologic processes perturbed by PoC chemicals. The first version of the KB will focus on the PoC chemicals and populated mostly with their molecular activity data from ToxCast^{TM}, and cellular and tissue level outcomes from ToxRefDB and the literature.
- Deploy KB visualization tool for client interaction. Access to the KB will be provided using open source tools for biological data analysis.
- Simulate of liver lesions for alternative MOA/toxicity pathways. The prototype of the lesion simulator implementing the main MOA for hepatocarcinogenesis.

FY10

- Evaluate simulator using PoC chemicals and ToxCast data to predict outcomes.
- Quarterly update of v-Liver KB
- v-Liver KB inference tool for analyzing MOA for new chemicals/mixtures
- Extend v-Liver Simulator to liver and integrate with PBPK model

FY11

- Evaluate impact of genomic variation on cellular responses and lesion formation
- Evaluate v-Liver for simulating human pathology outcomes using clinical data

Most milestones will also include manuscript submissions describing the computational methods and their biological/toxicological relevance.

v-Embryo:

- Literature-mining tools to index relevant facts about early eye development and concept model (ontology) to support this knowledge representation [2];

- Ocular gene network schema specified by gene-gene and gene-phenotype associations and subjected to dynamical network inference analysis; computer program of early eye development that reconstructs lens vesicle induction *in silico* using cell-based simulators and system-wiring diagrams of perturbation analysis of the computational (*in silico*) model with pathway-level data for normal and abnormal (toxicological) phenotypes *in vitro* and *in vivo*.

FY09

- Project plan and quality assurance plans for VT-KB and VT-SE
- Recruit: student contractor and postdoctoral fellow

- Manuscript: application of VT-KB to analyze ToxRefDB developmental toxicity studies
- Model: VT-KB based qualitative (structural) model of self-regulating ocular gene network
- Model: VT-SE based cell-based computational model of lens-retina induction
- Manuscript: ocular morphogenesis, gene network inference, analysis and modeling

FY10

- Project plan: extend lens-retina model to other stages and species
- Model: incorporate pathway data from ToxCast™, mESC and ZF embryos
- Manuscript: sensitivity analysis for key biological pathways
- Manuscript: analyze developmental trajectories and phenotypes in computational models
- Project plan: integrate with other morphogenetic models

FY11

- Manuscript: test model against predictions for pathway-based dose-response relationship
- Manuscript: uncertainty analysis of models for complex systems model: computer program of early eye development using rules-based architecture, cell-based simulators and systems-wiring diagrams