National Toxicology Program
Board of Scientific Counselors

December 9-10, 2009

National Institute of Environmental Health Sciences
Research Triangle Park, NC

Table of Contents

I. Attendees

Members in Attendance:
Tracie Bunton, Eicarte LLC
Edward Carney, The Dow Chemical Company
Russell Cattley, Amgen
David Eastmond, University of California
Janan T. Eppig, The Jackson Laboratory
Elaine Faustman, University of Washington
George Friedman-Jimenez, New York University School of Medicine
William Janzen, The University of North Carolina at Chapel Hill
Stephen Looney, Medical College of Georgia
Mitzi Nagarkatti, University of South Carolina School of Medicine
Raymond Novak, Wayne State University
Kenneth Portier, American Cancer Society (Chair)
Jim Riviere, North Carolina State University
Diane Robins, University of Michigan Medical School
Ruthann Rudel, Silent Spring Institute
Justin Teeguarden, Pacific Northwest National Laboratory

Members not in attendance:
Michael Pino, Sanofi-Aventis
James Sherley, Boston Biomedical Research Institute
Gina Solomon, Natural Resources Defense Council

Ad hoc:
David Threadgill, North Carolina State University

National Institute of Environmental Health Sciences Staff

Scott Auerbach
Patrick Barbour
Mamta Behl
Linda Birnbaum
Jack Bishop
Susan Borghoff
John Bucher
Rajendra Chhabra
Bradley Collins
Michael Cunningham
Helen Cunny
June Dunnick
Paul Foster
John French
Mary Gant
Dori Germolec
Deborah McCarley
Sheila Newton
Veronica Robinson
Michael Sanders
William Schrader
Thaddeus Schug
Barbara Shane
Keith Shockley
Robert Sills
Cynthia Smith
Diane Spencer
William Stokes
Inok Surh
Christina Teng
Kristina Thayer
Raymond Tice

Wanda Holliday
Michelle Hooth
Gloria Jahnke
Frank Johnson
Patrick Kirby
Grace Kissling
Sandy Lange
Ruth Lunn
Robin Mackar
David Malarkey
Scott Masten
Molly Vallant
Suramya Waidyanatha
Nigel Walker
Lori White
Kristine Witt
Mary Wolfe
Michael Wyde

Other Federal Agency Staff
Paul Howard, Food and Drug Administration (FDA)
Mark Toraason, National Institute for Occupational Safety and Health (NIOSH)

Public
Andrew Ballard, BNA, Inc.
James Blake, RTI International
Donna Browning, RTI International
Reshan Fernando, RTI International
Charles Hebert, Southern Research Institute
Liz Hill, RTI International
Cheryl Hobbs, ILS, Inc.
Marc Jackson, ILS, Inc.
Jeffrey Johnson, ILS, Inc.
Joseph Manuppello, PeTA
Glenda Moses, ILS, Inc.
Catherine Price, RTI International
Ivyn Rusyn, University of North Carolina at Chapel Hill
Michael Waters, ILS, Inc.

II. Introductions and Welcome

The National Toxicology Program (NTP) Board of Scientific Counselors (BSC) met on December 9-10, 2009, in Rodbell Auditorium, National Institute of Environmental Health Sciences (NIEHS), Research Triangle Park, North Carolina. Dr. Kenneth Portier served as chair. He welcomed everyone to the meeting and asked the BSC members and attendees to introduce themselves. Dr. Barbara Shane made a few announcements and read the conflict of interest statement. She noted that the *ad hoc* reviewer would not vote and no conflicts of interest were identified.

III. Report of the NIEHS/NTP Director

A. Presentation

Dr. Linda Birnbaum, Director of NIEHS and NTP, welcomed the BSC members and expressed gratitude for their service. She presented certificates of service to the retiring members, Drs. Friedman-Jiménez, Kenneth Portier, Jim Riviere, and Diane Robins.

Dr. Birnbaum briefly updated the BSC on the status leadership positions at the NIEHS, which include Scientific Director; Director Division of Extramural Research and Training; Deputy Director, NIEHS; Clinical Director; Director, NIEHS Ethics Office; Education Director; and four staff positions in the NIEHS Bethesda office. She highlighted some recent invited talks and events. These include:

- testimony at the Senate Oversight Hearing on the Toxic Substances Control Act (TSCA)
- a briefing on the NIEHS Vision to the NIH Director's Advisory Council,
- closing comments and moderated press conference at the Public Health Impacts of Reducing Greenhouse Gas Emissions workshop
- speaker and session chair at Energy Impacts on Human Health at the White House Stakeholder Briefing on Climate Change
- plenary speaker at Breast Cancer and the Environment Conference
- speaker at a Sausalito Community Forum
- keynote speaker at Moving Upstream: Thyroid Disruption Workshop
- keynote speaker at Mammary Gland and Risk Assessment Workshop
- plenary speaker at 19th Annual International Society of Exposure Science Conference
- opening remarks at ILSI-HESI Workshop, Evaluating Epigenetic Changes: State of the Science; Milwaukee Town Hall Meeting
- keynote speaker at 2009 Rachel Carson Legacy Conference

Regarding the NIEHS budget, Dr. Birnbaum reported that the Senate had not yet passed its funding bill for FY2010; there is no final appropriation and NIEHS is on a Continuing Resolution. In both the House and Senate budgets, NIEHS has the largest percentage increase of any NIH Institute/Center, 4.9% in the House and 3.1% in the Senate. NIEHS priorities in the FY2010 President's Budget include cancer, autism, environmental health, endocrine disruption, safety of nanomaterials, alternative test evaluations, and health issues related to climate change. Dr. Birnbaum anticipated funding in the NTP for FY2010 to be similar to FY2009 levels. Dr. Collins, NIH Director, has a list of priorities that include empowering the biomedical research community, global health research, health care reform, health disparities, high throughput approaches, the microbiome, small molecule screening, stem cell research, and translational research. Dr. Birnbaum outlined priority areas of the Office of Management and Budget and the Office of Science and Technology Policy: economic recovery, job creation, and growth; innovative energy technologies; biomedical science and information technology; technology to protect our troops, citizens, and national interests; and a number of cross cutting themes.

Dr. Birnbaum reminded the BSC that the scope of NTP's activities focuses on two broad areas: (1) research and testing and (2) analysis activities. She is working across NIEHS to foster ways in which the Institute as a whole can better work together for research planning and development of trans-NIEHS research priorities. The NTP's combination of research and testing and analysis activities covers a broad array of topic areas including children's health, herbals and dietary supplements, occupational exposures, cell phone radiation, phototoxicity, and nanoscale materials. There is a new and renewed emphasis for the NTP in many areas including coordinating toxicity testing across the Federal government, developing new methodologies for efficient and thorough toxicological assessments, establishing additional capacity for dosing during the perinatal period, increasing understanding of exposure-response relationships and issues of dosimetry, developing appropriate safety testing approaches for nanomaterials, integrating results from new "data rich" techniques (i.e., genomics, high through-put screening) with traditional toxicology data to provide public health context, and providing guidance for the proper utilization of new types of information in hazard identification and characterization.

Dr. Birnbaum reported on NIEHS American Recovery and Reinvestment Act (ARRA) opportunities. The NIEHS received approximately $190M, of which $168M was used for environmental sciences, $19.4M for Superfund, and $11.4M for NIH/OD contributions and co-funds. By the end of FY2009, the NIEHS allocated 95% of NIEHS ARRA money to 346 grants. NIH received 598 proposals for Challenge Grant awards and funded 38 of them including those in two signature areas, health effects of bisphenol A and nanomaterials safety.

IV. NIEHS/NTP Update

A. Presentation

Dr. John Bucher, Associate Director of NTP, updated the BSC on staff changes at the NTP in the Program Office, the Center for the Evaluation of Risks to Reproduction (CERHR), the Program Operations Branch Chemical Effects in Biological Systems (CEBS) database, the new NTP Laboratories Branch, the Biomolecular Screening Branch (BSB), and the Toxicology Branch postdoctoral programs and Reproduction and Development Group. The NTP is actively searching for additional new staff in pathology, toxicology, and the Program Office. He announced the retirement of Dr. Barbara Shane. Dr. Shane said it had been a great pleasure working with the BSC and she would miss her colleagues at the NTP.

Dr. Bucher said the Report on Carcinogens (RoC) Center held a formaldehyde expert panel meeting in November 2009, which concluded expert panel evaluation of substances under consideration for the 12th RoC. The Technical Reports Review Subcommittee (TRRS) met in November; a report to the BSC would occur at a later meeting. The next TRRS meeting, reviewing acrylamide, retinyl palmitate/retinoic acid phototoxicity, AIDS therapeutics, kava kava extract, alpha-beta thujone, methyl trans-styryl ketone, styrene acrylonitrile trimer, and pyrogallol, would be held in late 2010.

Dr. Bucher described the launch of the new NTP Laboratories Branch. The laboratories would be available to all NTP staff and would focus work in two areas. The first would be to provide the capability to address immediate needs of the NTP for short-term studies and enable the NTP to better design research studies that would then be done by contract laboratories. The second area would be to develop better methods to study the developmental origins of adult diseases. The NTP Laboratories also house the *Caenorhabditis elegans* (*C. elegans*) WormTox Program for medium throughput screening of chemicals. The study design, review, and evaluation processes for the NTP Laboratories will be identical to those for all other NTP projects to ensure the highest scientific quality and relevance to the NTP mission.

Dr. Bucher noted that the BSB is headed by Dr. Raymond Tice and Dr. Jonathan Freedman providing guidance for its WormTox Screening Program. A major initiative of the BSB is participation in Tox21 with EPA and the NIH Chemical Genomics Center (NCGC). Dr. Bucher described the four working groups comprising the Tox21 Community: Pathways/Assays, Compounds, Bioinformatics, and Targeted Testing. Compound selection has been the current focus of Tox21, with the goal of creating a library of 10,000 chemicals. The chemical list would be approximately 1/3 drugs, 1/3 chemicals selected by the EPA, and 1/3 chemicals selected by the NTP. Identity, stability, and purity information is being collected on all chemicals. Tox21 would establish a small library of mixtures and a library of water-soluble compounds. The Bioinformatics Group is evaluating patterns of response and their relationship to adverse health outcomes in experimental animals and humans, evaluating the consistency of the response within and across assays/endpoints, and making all data publicly accessible. The Targeted Testing Group is prioritizing substances for more complex testing including the use of alternative assay platforms. The Phase I HTS assays currently being used at the NCGC include cell viability, apoptosis, pathways, DNA damage, epigenetics, inter-individual variation in chemical response, and nuclear receptors.

The WormTox Group has completed testing of 20 chemicals in the ToxCast™ Program in a growth assay; analyses of those data are in progress. The NTP is collaborating with Hemogenix to utilize stem cells to test for immunotoxicity of 25 compounds. The NTP is conducting pilot toxicogenomic and proteomics studies on formalin-fixed tissues from the NTP Archives and is also conducting a sequencing pilot study on rat tissues and tumors.

A significant advance in the Bioinformatics Group is development of a statistical framework for analyzing quantitative high throughput screening Data HTS (qHTS), which will be presented at the 2010 Society of Toxicology (SOT) meeting in March 2010. NIEHS recently announced several requests for Small Business Innovation Research/Small Business Technology Transfer (SBIR/STTR) proposals, which are anticipated to help with NTP HTS efforts.

Dr. Tice gave the plenary lecture at the 7th World Congress on Animal Alternatives in Rome, August 26 - September 3, 2009. Due to the early success of Tox21, there is

interest by the US Food and Drug Administration (FDA), Agency for Toxic Substances and Disease Registry (ATSDR), and European Commission's Institute of Health and Consumer Protection to partner with Tox21. The Board of Scientific Counselors for EPA's National Center for Computational Toxicology's (NCCT) met recently and strongly supports action by EPA to make the NCCT permanent.

Dr. Bucher reiterated some of the areas of emphasis and challenges facing the NTP, which include high content data, genomics, Toxicology in the 21st Century, new criteria for non-cancer endpoints, and societal expectations. He highlighted the progress the NTP has made in addressing these issues, such as new hires in CERHR and new processes, products, and scope for CERHR; potential streamlining of the RoC review process; new partners in Tox21; targeted testing including testing herbals/dietary supplements in coordination with FDA; and the International Cooperation on Alternative Test Methods Memorandum of Cooperation. Dr. Bucher explained that the NTP is reaching out to other groups with the goal of improved public understanding of our mission and products.

B. BSC Discussion

Ms. Rudel said she was pleased and encouraged by the information Dr. Bucher presented. Dr. Portier thanked Dr. Shane for her outstanding management of past BSC and TRRS meetings.

V. Center for the Evaluation of Risks to Human Reproduction

A. Evaluation Process

Dr. Mary Wolfe, NTP Deputy Program Director for Policy, pointed out the two types of CERHR evaluations:
(1) Level of concern (LoC) assessments that determine potential adverse effects on reproduction and/or development caused by environmental substances at current human exposures. These follow an established, formal process and result in a 5-level scale of concern (serious concern to negligible concern) that is published as the NTP Brief in a NTP monograph (e.g., BPA).
(2) State of the science evaluations, e.g., CERHR workshop, Thyroid Toxicants: Assessing Reproductive Health Effects (April 2003).

Dr. Wolfe described the current CERHR LoC evaluation process, which has three parts:
(1) Nomination and selection of candidate substances. An interagency group reviews the nominations, which can come from anyone. In proposing substances for evaluation, the CERHR considers production volume and use, extent of human exposure, degree of public concern, and extent of data on reproductive and developmental toxicity studies. The CERHR solicits public comment on proposed substances and the NTP brings proposed evaluations to the BSC. The NTP then considers all inputs and selects candidate substances for evaluation.
(2) Scientific evaluation of candidate substances. The NTP assembles the relevant scientific literature on the substance (previously with assistance from a contractor) and then convenes an independent external scientific panel. Members of the panel are

selected to provide relevant scientific expertise in accordance with Federal Advisory Committee Act (FACA) and applicable federal policies and regulations. The panel reviews the literature and public comments and develops a draft report that includes their assessment of the strengths/weaknesses of the studies and their utility to the current evaluation. The draft report (without the panel's LoC conclusions) goes out for public comment. The panel then convenes in a public meeting to receive additional public comment, reach their LoC conclusions for adverse reproductive/developmental effects associated with current human exposures, and identify data gaps and research needs. After the meeting, the NTP releases the final expert panel report for public comment. The end product of the scientific evaluation is development of the draft NTP Brief. The final expert panel report has been the primary document used by the NTP to prepare the draft NTP Brief, although the NTP continues to monitor the literature to identify new, relevant studies. The NTP Brief is the document that gives the NTP's LoC conclusions for the substance at current human exposures. Prior to moving to part 3, the NTP seeks interagency review of the draft NTP Brief.
(3) Peer review and release of NTP monograph: This part of the process was added in 2004 in response to guidance provided in the Office of Management and Budget Final Information Quality Bulletin for Peer Review. The draft NTP Brief is released for public comment and peer reviewed by ad hoc experts or the BSC. The NTP finalizes the NTP Brief based on peer review comments and public comments. In the final step, the CERHR compiles the NTP Monograph, which consists of the final NTP Brief, expert panel report, and public comments on final expert panel report, and releases it to the public.

The CERHR is considering revising the evaluation process to allow more flexibility to tailor the scientific evaluation of candidate substances to fit the "evaluation's needs." Moving away from this "one-path" format would also allow integration of new CERHR staff into the evaluation process and potentially shorten the time needed for an evaluation, which can currently take up to 25 months. Dr. Wolfe explained that the proposed revised process would still remain a 3-part process; retain the opportunities for scientific input, public engagement at multiple points, interagency input, and peer review; and produce the NTP monograph. In part 1, the proposed changes would allow an opportunity for the BSC and public to review the proposed evaluation approach in a concept document. Part 2 would no longer have a set format; the external scientific input, public input, and interagency input would not occur in any set order, but would depend on programmatic needs. The mechanism(s) used to obtain external advice and address scientific issues would also depend upon programmatic needs: topic, nature and extent of the literature, degree of scientific complexity, and public interest. Types of external scientific input might include: external ad hoc participants like the current expert panels, invited presentations by experts in the field at a public meeting, or a workshop to discuss a topic and get broad scientific and public thought on a current theory. Types of public input might include: written and oral comments as the NTP currently does, in concert with release of a draft document or public meeting, but also holding public listening sessions where the public is given 20-30 minutes to present views on a topic or draft. The NTP might also do this type of solicitation through written comments on specific questions. The NTP would continue to seek interagency input during the

scientific evaluation and the end product of the scientific evaluation is development of the NTP's opinion as the draft NTP Brief. Part 3 would not change; the NTP would continue to choose the most appropriate mechanism for peer review and it would be followed by finalization and release of the NTP Monograph (NTP Brief and literature review).

Dr. Wolfe presented two examples of the new process using previously examined chemicals that vary in complexity. Propylene glycol, evaluated in 2003/2004, had a small database, no human data, and low public interest. For this chemical, CERHR's proposed process likely would have included a public listening session and no independent external scientific input. BPA, evaluated in 2006/2008, had a large database, high scientific complexity, and high public interest. The proposed scientific evaluation for BPA, using the revised CERHR process, likely would have included an independent, external scientific panel, a listening session and public comment on the draft and final expert panel reports.

Dr. Wolfe said the NTP believes this revision to the CERHR process, to enhance scientific evaluation of candidate substances, is a positive step for the program and welcomed comments on the revised process and suggestions to make it better.

B. BSC Questions

Dr. Teeguarden asked for clarification of the difference between the NTP Brief and the expert panel report and if it was possible to have public review of the panel report and NTP Brief at the same time to save a review step. Dr. Wolfe explained that the peer review is of the government opinion, not of the expert panel report. For complex evaluations, the NTP has sought input from external experts to provide guidance for the NTP's conclusions. The need for an expert panel is dependent upon the nature of the evaluation. Dr. Bucher added that the expert panel report represents advice to the NTP. In the past, the NTP has solicited public comment on the advice before it was taken under consideration. The final monograph consists of the expert panel report and NTP Brief. With the new process, what were two distinct steps is now merged into one step, which is creation of the NTP opinion. There is now a flexible way of getting public and expert input and developing the opinion.

Dr. Eastmond supported streamlining the process, but was concerned about bypassing independent expert evaluation of a substance. He said a formal panel may not be necessary, but the draft document should have a review by letter or teleconference to get additional perspective.

Dr. Carney agreed, stating that the expert panel is key to an evaluation's success. He supported having flexibility on public input and questioned the need for a rigorous CERHR-like process for low-importance chemicals. Dr. Birnbaum said the evaluation type depends on BSC's opinion. Dr. Wolfe said the first step in the process is for the NTP to prepare an approach that the BSC would evaluate.

Dr. Friedman-Jiménez asked about expert panel selection and balance and if clinicians are included. Dr. Wolfe said the NTP identifies experts in reproductive and development toxicity, exposure assessment, industrial hygiene, and any other appropriate scientific discipline. The experts are reviewed for potential conflicts of interest and bias toward the substance under review. Panels are also balanced for ethnicity, gender, and geographical area. Panels in the past have included pediatricians and obstetrician/gynecologists.

Ms. Rudel asked about the opportunities for peer review before the brief is prepared. Dr. Wolfe explained that peer review occurs after the brief is completed. Ms. Rudel supported combining the reviews, but expressed concern about having the appropriate expertise for the review. Dr. Wolfe said the BSC would peer review the draft brief or the NTP would convene an external panel, whatever is most appropriate. Dr. Bucher added that concepts for selecting projects to go through the CERHR evaluation are brought before the BSC, but now would also provide input on the suitable process to be used to evaluate the substance. Ms. Rudel suggested using the broader term "issue" in lieu of the word "substance" to be more inclusive.

Dr. Howard concurred that there is value in having many opinions during the review process, but considered the initial concept presentation before the BSC as adequate to guide the review. He asked about potential ad hoc expertise on the BSC to evaluate a concept. Dr. Wolfe said the NTP reviews the expertise on BSC for a given substance and obtains additional expertise if needed. Dr. Howard said low public interest does not necessarily indicate low public health agency interest.

Dr. Portier said for the BPA evaluation there was a definite need for an expert panel due to the large body of literature; however, the propylene glycol evaluation could have easily been done without an expert panel. He supported the revised process.

C. CERHR Update and Evaluation Concept: State of the Science Evaluation on Environmental Exposures and Diabetes and Obesity

Dr. Kristina Thayer, CERHR Acting Director, said, regarding the previous discussion, that she doubted there would be a case where the NTP would issue a draft opinion without soliciting input from outside independent experts. In instances where an expert panel is not convened, the NTP would most likely use an *ad hoc* special emphasis panel for peer review. Dr. Thayer reported that CERHR hired 4 new staff with expertise in reproductive and developmental biology, endocrinology, immunology, epidemiology, and risk assessment. She said CERHR staff would work with the Office of Liaison, Policy and Review to ensure that evaluation processes meets its goals and that each evaluation is consistent with Federal Advisory Committee Act implementing regulations.

Dr. Thayer provided an update on the soy infant formula evaluation identifying important timepoints: (1) October 2009, CERHR released the draft expert panel report; (2) December 2009, the expert panel would meet and develop their LoC conclusions; (3) January 2010, CERHR would release the final expert panel report for public comment; (4) release of the draft NTP Brief on Soy Infant Formula for public comment in March

2010; and (5) May 10, 2010, the BSC would peer review of draft NTP Brief and CERHR hopes to release the final NTP Monograph on Soy Infant Formula in the summer 2010.

Dr. Thayer presented the concept proposal for the State of the Science Assessment of Environmental Exposures and Diabetes/Obesity. The justification for this assessment is that diabetes and obesity are major threats to human health. Forty percent of people over the age of 20 have diabetes or pre-diabetes and ~70% of type 2 diabetes risk can be attributed to overweight/obesity. Diabetes and obesity are often discussed in the context of developmental origins of adult disease or prenatal programming. There is growing scientific and public interest on this issue. An association of diabetes with organohalogens/dioxin, arsenic, and BPA has been identified. The assessment builds on a previous NTP workshop *Biomarkers for Toxicology Studies* held in 2006.

The goals and expected outcomes of the assessment are: (1) complete a critical assessment of the current literature and (2) focus future research directions. The proposed approach is to (1) present the concept to the BSC, (2) obtain public comment on the literature review document, (3) convene a public workshop in early fall 2010, and (4) prepare the NTP Monograph in late fall 2010. The literature screen for the assessment was an iterative process using a combination of health- and exposure-based medical subject heading terms. The nature of literature varies from exposure to exposure, i.e., number of publications and proportion of human/animal/*in vitro*, which indicates that the type of conclusions reached for each type of exposure would vary. CERHR's approach to assessing obesogens is using a very targeted search. It was necessary to consider the consistency of effects on body weight following developmental exposure and to recognize that this topic is more complex than effects on just body weight. For example, Dr. Thayer noted *in utero* exposure to tributyltin increases adipose mass but not body weight in adult mice (Grun *et al.,* 2006)[1].

The first stage of the assessment would require expertise in particular exposures as well as expertise in the primary health conditions. Expertise for the second stage of assessment would overlap with first, but would be supplemented with other individuals to generate both general research suggestions as well as those specific to NIEHS/NTP.

Dr. Thayer presented the charge to BSC: Review and comment on the proposed CERHR evaluation concept and determine whether the evaluation is an appropriate use of NTP program resources.

D. BSC Discussion

Dr. Riviere responded that the assessment is an appropriate evaluation. Obesity is a major issue for the Food Nutrition Board. A number of potential factors in metabolic syndrome are based on early exposures, but there are no data to tie things together.

[1] Grun, F., Watanabe, H., Zamanian, Z., Maeda, L., Arima, K., Cubacha, R., Gardiner, D. M., Kanno, J., Iguchi, T., and Blumberg, B. (2006). Endocrine-disrupting organotin compounds are potent inducers of adipogenesis in vertebrates. *Mol Endocrinol* **20**, 2141-55.

He suggested contacting the Food Nutrition Board for suggestions on experts to include.

Dr. Nagarkatti said this is a very significant area of research and questioned how CERHR would prioritize chemicals that are to be evaluated. Dr. Thayer responded that discussions with the experts at the workshop would help triage the chemicals. Dr. Birnbaum suggested replacing the term "exposure" with "agents" or "substances." Dr. Teeguarden clarified that the 30% of diabetes that is not related to obesity is related to chemical exposures, nutrition, and life style.

Dr. Thayer noted in response to Dr. Eastmond that there would be sequential workshops in the evaluation process. Dr. Thayer responded to Dr. Carney that the workshops are not structured to develop a LoC conclusion or a labeling of a substance as an obesogen/diabetic agent. However, the workshop report might conclude that there should be a separate evaluation activity to determine a LoC for a particular substance.

Dr. Novak said this is a large issue with multiple perspectives. He recommended devoting efforts at the workshop to address mechanisms, because both oxidative stress and basal inflammation underlie the conversion from obesity to type 2 diabetes. He mentioned the defined effects on signaling components downstream from receptor tyrosine kinase through the kinases that regulate insulin resistance. Dr. Thayer said CERHR planned to have experts in mechanisms on the first subpanels. The CERHR would also include tables in the documents that clearly show the mechanisms and patterns involved.

Dr. Carney, the first lead discussant, commented that the rationale for the assessment is strong and the project aligns with CERHR's mission to strengthen the science base in toxicology. There is considerable interest in the topic and broad societal implications. He suggested changing the title to clarify that developmental exposures are being evaluated. The research area is quite new, the number of laboratories doing the research is small, the area of study is difficult, a multifactorial disease etiology is postulated, and newer endpoints are being evaluated. Given all of these factors, three years is not much time to amass a robust amount of data that can be interpreted in any broad sense. He agreed that the topic is worthy of a CERHR workshop, but the timing could be too early. Unlike this topic, the state of the science evaluation of thyroid toxicants and reproductive health, published by the CERHR in 2003, had a much deeper scientific literature upon which to draw. Dr. Carney cautioned against doing hazard assessments on individual chemicals. If the assessment is confined to a just a workshop to guide future research, he would assign a high priority to the project. He suggested waiting for more data to amass, or to broaden the scope beyond obesity and diabetes and make it more generic to developmental origins of adult disease.

Ms. Rudel, the second lead discussant, said the rationale and need for this assessment were clearly articulated and the activity has high merit and is proactive. This is an emerging area of research with a great deal of activity, a high level of interest and public

health implications, and unresolved science. The project would support increased efficiency in research and assist the NTP in prioritizing endpoints and potentially provide input into HTS studies. It could synthesize findings across multiple chemicals and disciplines and highlight key questions and areas where new knowledge is taking shape. Ms. Rudel considered the scope to be on target and concurred with the three key focus areas. She suggested having the subpanels focused on particular outcomes, disease markers, or mechanisms. Some of their work should be to synthesize across chemicals, and some simply to characterize pathways to disease, relevance of various markers, etc. There are outstanding questions on biomarkers of effect and the pathway to outcomes. Ms. Rudel urged CERHR to have epidemiologists and toxicologists working together on the subpanels. She questioned why CERHR would do this project, since it isn't related to reproduction. She suggested using words such as "assessment" and "evaluation" in lieu of words like "conclusions" and "findings," which imply a clarity that isn't consistent with the state of the science.

Dr. Thayer said there is a large literature on arsenic, organochlorines, dioxin, and genistein in both humans and animals. Dr. Carney said the quality of the studies must be considered, and questioned the relevance of some studies in terms of route of exposure and high doses. Dr. Thayer said a reason for doing this project is evaluate the developmental origins of adult disease, but studies in the literature often used adult exposures.

Dr. Birnbaum suggested thinking more broadly. She said there is clearly an epidemic of obesity and diabetes in children today. However, when studies are being done in adults, and there is no information on prior history, it is hard to link exposures 30 to 50 years ago to adult disease. In planning the workshop, CERHR should focus the questions being asked. The project should be integrated with the larger cross-NIH efforts on diabetes and obesity as well as with the NIEHS intramural and extramural efforts. Dr. Thayer said the first round of subpanels might come from the NIH extramural community. Dr. Birnbaum considered the workshop a good idea to assess what is known, but cautioned against limiting the study to traditional models.

Dr. Teeguarden said scientists continue to improve their ability to assess response and understand mechanisms, but struggle with interpretation of the data in the context of real world human exposure. He asked if assessment of the studies would include examination of the relationship between the doses and target tissue exposure and if comparisons would be made to human exposures. Dr. Thayer said the literature is not good enough to make those comparisons. The goal of the workshop is not to compare animal and human exposures, but rather to assess the current findings and determine what should be done next. Dr. Teeguarden said dose should be considered in terms of relevance to mechanism. Dr. Thayer said, using arsenic as an example, that CERHR would prepare summary tables listing the administered dose or concentration for an *in vitro* study. Typically the doses are high compared to human exposures. Dr. Teeguarden said the EPA has an effort underway to develop pharmacokinetic models for arsenic. He suggested including dose comparisons in the evaluation or requesting those studies in the data gap analysis. Dr. Thayer said for arsenic and organochlorines,

there are more directly relevant studies in humans than in animals. Dr. Bucher said biological plausibility would be a focus of the project. The NTP had a previous workshop on identifying biomarkers in animal studies that could be used to identify inflammation. The clear message from the workshop was that animal models are poor for identifying inflammation. Biomarkers used in humans are not always that good for animal studies. The development of research gaps and research recommendations feeds into the kinds of things the NTP laboratories are going to be focusing on as they put together research programs to develop better ways of analyzing animal studies for developmental effects that might result in chronic diseases. Dr. Cattley said animal models create a lot of concern in terms of whether data are generated that address real world issues. He said one valuable outcome of the workshop would be a meeting of minds between those who characterize and treat the human disease and those looking forward and developing new model systems.

Dr. Eastmond approved of the concept and considered it important, but expressed concern about the broad focus and suggested it might be narrowed. Because many of the studies are very new, there needs to be time for replication of studies to ensure they are reproducible. He urged caution and suggested allowing three to four years to see replication or refutation of the studies.

Ms. Rudel questioned whether the arsenic literature is complete enough to make a strength of evidence conclusion. Dr. Thayer said studies on areas of high arsenic in drinking water, such as Bangladesh and Taiwan, report a consistent association, but the low to moderate exposures studies have prompted serious debate. Mr. Janzen asked whether the full review process would be used if the expert panel determines a clear assessment of risk. Dr. Thayer said it would become a separate evaluation activity that would be presented to BSC.

Dr. Bunton agreed that the evaluation is too early, though there is a huge database. She suggested the workshop be used as a way to focus information rather than trying to make any firm conclusions. CERHR could then guide how this research area would move forward. Dr. Thayer said that is an overarching goal.

VI. Use of Contracts in the NTP Testing Program

A. Overview of the Use of Contracts in the NTP Testing Program

Dr. Cynthia Smith, NIEHS/NTP, provided an overview of the use of contracts in the NTP testing program. The NTP makes extensive use of contracted resources to provide necessary facilities, additional personnel, and abilities to accommodate changes in the NTP's direction, new capabilities when needed, flexibility in access to specific expertise, accountability, and specific deliverables. Each of the contracts is directed by NTP personnel, called Contacting Officer's Technical Representative (COTR), with the appropriate expertise. The statement of work (SOW), which is approved by management and contract officials, specifies the scope of the contract activities. The COTR is responsible for directing and approving work conducted under a contract.

Dr. Smith provided an example of a typical 90-day general toxicity study and described the phases of the study that are initiated with a substance nomination: (1) study planning and design, (2) study conduct, (3) results and data review, and (4) report preparation and production. These culminate in a published report.

The use of contracted resources allows: (1) the NTP to concentrate in-house expertise in study design and oversight, (2) checks and balances throughout the testing program, and (3) flexibility to change the type and amount of resources with changes in scientific approach, policy, and budget. As a result, the NTP can be more responsive to arising public health needs while maintaining data quality and scientific rigor.

B. BSC Questions

Dr. Bunton asked how the NTP works with contractors on issues that are rapidly emerging. Dr. Smith said many of the larger contractors have dedicated NTP contract personnel that allow specific people and equipment to be used for NTP projects.

Dr. Friedman-Jiménez said he could see huge benefits to using hundreds of contractors as the NTP diversifies, but expressed concern about potential loss of cross-fertilization among scientists in different fields within NTP. Dr. Smith said lines of communication go both ways with contractors. The NTP listens to the experts that are hired and makes the decisions, but many of the contractors have extreme expertise and provide valuable feedback, so cross-fertilization does occur with the contractors.

Mr. Janzen asked if all the contractors were U.S. based. Dr. Smith said all major contract work has been done in the United States, with only a few small subcontracts being done by foreign contractors.

Dr. Nagarkatti asked about the number of personnel in the NTP and at what point statisticians become involved in study designs. Dr. Bucher responded that approximately 100 people in NIEHS are involved in the NTP contract studies. Others from the National Center for Toxicological Research (NCTR) and the National Institute for Occupational Safety and Health (NIOSH) are also involved in NTP contract work. Dr. Smith said statisticians are involved very early in the project-planning stages, especially in determining how large a study should be.

Ms. JoAnn Lewis, NIEHS Office of Acquisitions, briefly outlined the guidelines for the BSC regarding the discussion of contract concepts. She asked the BSC to review two concepts--analytical chemistry support and laboratory studies for the NTP to evaluate toxicity following early life exposure to chemicals--for their overall value and for their scientific relevance to fulfill the program's goal of protecting public health. They should consider the availability of technology to achieve the required goals, adequacy of the methodology to be used to perform the activity, the scientific or clinical uses of the anticipated data, and scientific, technical, and programmatic significance of the proposed activities. The discussion should be limited to a review of the general purpose, scope, goal, and optional approaches to pursue the overall objectives. The meeting would be closed to the public should discussions turn to the development or

selection of the details of the project such as specific technical approaches, protocol, statement of work, data format, or product specifications. A meeting is closed to protect free exchange of the advisory group members' opinions and avoid premature release of the details of the proposed contract or request for proposal.

C. Contract Concept Review: Analytical Chemistry Support for the NTP

Dr. Smith provided a brief history of the analytical chemistry support contracts, which have been used since the inception of the NTP. The current contracts have been in place in their present form (with updating) for nearly 15 years. Support activities include specific tasks in: logistics (e.g., chemical procurement and shipping), chemical characterization, dose formulation development, biological sample analysis, and toxicokinetics studies with unlabeled compounds. She listed the types of analytical techniques used for analytical testing of NTP chemicals. The program areas supported by the analytical support contract include reproductive toxicity, developmental toxicity, carcinogenicity, and general toxicity. In addition, special support is provided for certain interagency projects, NIEHS Division of Intramural Research projects, and the HTS Program.

Analytical chemistry using accurate and robust methods for the characterization of test articles, formulation and analysis of dosage preparations, and measurement of test articles and metabolites in biological matrices is critical to the scientific integrity of NTP studies. Furthermore, logistical activities (purchasing test articles and shipping test articles, formulations, and samples) coordinated with other NTP contract laboratories are vital to the productivity of the program as a whole.

D. BSC Discussion

Dr. Nagarkatti asked if testing of chemicals includes storage stability. Dr. Smith said stability testing, including forced degradation testing, is included in test article characterization. Mr. Janzen expressed full support for the contract, stating that it fits with the NTP's activities and makes sense logistically and from a business standpoint. He moved to approve the contract concept. Dr. Novak seconded the motion and it passed unanimously (14 yes, 0 no, 0 abstentions).

E. Contract Concept Review: Laboratory Studies for the NTP to Evaluate Toxicity following Early Life Exposure to Chemicals

Dr. Helen Cunny, NIEHS/NTP, provided background for the contract concept for laboratory studies for the NTP to evaluate toxicity following early life exposure to chemicals. The NTP actively investigates adverse reproductive and developmental effects in rodents as part of its broad mandate to characterize the toxicity of agents of public health concern. Currently one in three nominations requires reproductive and developmental toxicity testing and there is now a backlog of 30 compounds for reproductive and developmental studies. This proposed contract would allow: (1) the NTP to increase its capacity to conduct these studies and (2) highly flexible study designs to evaluate effects of early life exposures over multiple life stages.

There is a need for reproductive and developmental testing because the adverse effects of chemical exposures during early life stages of gestation and the perinatal period continue to be an area of concern for the scientific community and the general public. The etiology is largely unknown for many problems related to reproduction and development such as difficulty achieving pregnancy, pregnancies not going to term, birth defects, invasive childhood cancers, and school-aged children diagnosed with Attention Deficit and Hyperactivity Disorder. Fetal and early postnatal life constitutes a vulnerable time for development; disturbing critical developmental processes can lead to functional deficits and increased disease risks later in life. A flexible testing paradigm is needed to allow the NTP to evaluate multiple endpoints using a variety of dosing schedules. The NTP is planning a "womb to tomb" assessment paradigm in which there is flexibility in the core exposure period (depending on test article-specific needs) and the ability to conduct single generation or multi-generation study designs. Flexible endpoints include assessment for: (1) teratology, (2) heritable genetic defects, (3) developmental neurotoxicity, (4) late life consequences (e.g., cancer), and (5) reproductive effects such as mating behavior, fertility, and fecundity.

Dr. Cunny said the NTP has a need for increased flexibility in study designs and increased capacity to conduct studies to evaluate adverse effects of exposure to chemicals during early life stages. The NTP seeks the BSC's approval to use a contract mechanism for this work

F. BSC Discussion

Dr. Cattley said there is clearly a need for the NTP to conduct early lifetime exposures and there is clearly public health significance. The effort is well justified and appropriate. Some points to consider regarding feasibility include: (1) balancing the need for Good Laboratory Practice (GLP) laboratories to conduct the studies with having the necessary flexibility for the studies with special endpoints and study designs, (2) whether the rodent model is sufficient to meet the needs of the NTP to evaluate early exposures, and (3) whether the contract concept would allow for mechanism of action evaluations. He encouraged the BSC to approve the contract concept.

Dr. Carney concurred with Dr. Cattley and approved of the concept because it allows more efficient use of animals. He noted that some contract laboratories might not have all the capabilities to complete these large, complex studies. He asked whether the NTP audits animal welfare practices at contract laboratories.

Dr. Cunny said GLP laboratories are preferred for contract work where possible. The NTP would incorporate into the solicitation the use of animals other than rodents if they were needed. The NTP has the capability to collect tissue samples for mechanistic studies; the mechanism of action work may then be done at other specialized laboratories. Dr. Birnbaum added that tissues from NTP studies are always available in the NTP Archives for other researches to use to address specific questions. The NTP is working to design studies so outside investigators can work cooperatively with NTP to do more mechanistic work.

Dr. Cunny said finding laboratories that can do large, complex studies is a challenge; the NTP may need to award more than one contract to obtain the flexibility and capabilities for the projects. The NTP works closely with the contract laboratories to ensure adherence to the best animal welfare practices.

Dr. Teeguarden said it is important to consider not just exposure, but tissue dose. He suggested assessment of internal dosimetry (e.g., of amniotic fluid and blood) in NTP studies, concomitant with collecting response data. Dr. Cunny responded that collection of body burden data would be included in the design of the studies.

Dr. Friedman-Jiménez agreed with the contract concept and observed that there has been a large increase in number and types of studies the NTP is doing in the last decade. The increase in scope is not commensurate with an increase in funding. The contracting mechanism gives some increase in efficiency and flexibility, but ultimately there is a limit to doing more work with the same funding. He suggested requesting additional funding from Congress. Dr. Birnbaum explained that the NTP could only help Congress understand the exciting and essential work the NTP is doing. Additionally, BSC could assist by working with advocacy groups and congressional representatives to inform them of the NTP's work. She said NIEHS is anticipating a 2.6% increase in funding; much of the $30 million increase is targeted for extramural research, but some additional funds would go to the NTP. In Federal agencies and universities, budgets have not kept pace with inflation and due to the aging workforce, personnel costs are high. Dr. Birnbaum said economic times are difficult right now, so big improvements in the budget cannot be expected. The NTP is working to increase awareness on Capitol Hill of the work the NTP is doing. The NTP would soon become a separate division, which should help with its visibility.

Dr. Cattley moved to approve the contract concept. Dr. Friedman-Jiménez seconded the motion and it passed unanimously (14 yes, 0 no, 0 abstentions).

VII. Review of Host Susceptibility Program

Four BSC members; Janan Eppig, Stephen Looney, Raymond Novak, and Diane Robins, and the *ad hoc* reviewer, David Threadgill, served as primary reviewers for the BSC review. The BSC and *ad hoc* comments are compiled into the attached report, Review of the NTP Host Susceptibility Branch Research and Testing Program, which is appended to the minutes.

A. Overview of the Host Susceptibility Branch (HSB) Research and Testing Program

Dr. John French, NIEHS/NTP, acting chief, HSB, provided an overview of the HSB and described the evolution of the NTP program. The HSB mission is to develop and test genetically modified and diverse animal models to determine (a) the range of their biological responses to toxic agents of public health importance, (b) the genetic and epigenetic bases for these biological responses, and (c) the mode or mechanistic basis for agent-specific toxicity in order to improve extrapolation across species. The primary

aim of research carried out by the HSB is to develop new laboratory models and testing protocols for hazard identification and risk assessment. The models are intended to quantitatively capture the range of response between individuals that correlate with individual human susceptibility to toxicity and disease. Individual genetic differences harbored within the human population are believed to be the basis for individual susceptibility to environmental stressors that cause toxicity, including idiosyncratic drug toxicities. The goal of the HSB is to model this genetic diversity in the human population using genetically diverse laboratory animals for *in vivo* and *in vitro* toxicology studies. The BSC was asked to review the eight projects in the HSB and then address the charge: Review and evaluate the research activities of the HSB related to the development of an NTP program in host susceptibility.

BSC Discussion
Dr. Teeguarden suggested highlighting "dose" in discussions of extrapolation across species in absorption, distribution, metabolism, and excretion (ADME) studies. Dr. French clarified that "dose" is included in environment (E) in the formula PHENOTYPE=[G + E + (G X E)] X T.

B. Background and Overview of the Perlegen Resequencing Project
Dr. Frank Johnson, NIEHS/NTP project leader, presented an overview of the NIEHS/NTP-Perlegen Resequencing Project. He explained that in 2004, the NIEHS/NTP embarked on a project to determine the genomic DNA sequence of 15 inbred strains of mice. The project was conducted under contract to Perlegen Sciences in Mountain View, California. The NTP is contemplating how to make best use of the findings to the benefit of toxicology, the testing of chemicals, and public health. Most toxicology testing utilizes very few genotypes, which tacitly assumes genotype makes little or no difference to test outcome. The study utilized 11 classical mouse strains and four wild-derived strains. The results of this project would help identify genes in mice that underlie susceptibility to adverse health conditions such as cancer and heart disease. The results can also help identify genetic factors responsible for variability in the response to toxic agents and help explain why some genotypes may be more susceptible than others to the harmful effects of exposure.

BSC Discussion
Overall, the BSC commended the NTP on the project and said the resequencing project's models reflect remarkably well what is present in human populations. This is a great resource that would serve the toxicology community and broader research community as a whole for years to come. These data have been incorporated into a number of databases that allow widespread access. This work has stimulated research in the broader scientific community, so rather than focusing on new sequencing projects, the HSB should focus on applying this information to the other HSB proposed projects.

Dr. Eastmond asked how many mice were evaluated when genotyping strains and how much variability there is within a strain. Dr. Johnson replied that the genotyping was a compilation of sequence across three to five 5 mice on non-overlapping regions; the

amount of variability is unknown. Dr. Nagarkatti asked how the NTP chose the 15 strains of mice. Dr. Johnson said seven of the strains were based on sales records at Jackson Laboratories; the remaining selections were attempts to cover the phylogenic tree.

Dr. Teeguarden asked about the timing of future projects and if information from the mouse strain models would guide the selection of additional strains and species and narrow the number that must be sequenced to understand variability in human populations. Dr. Johnson replied that prior studies are already providing some guidance at NIEHS and interesting variations have been found very early, but it is important not to get bogged down in a few phenotypes until the big picture is understood.

Dr. Eastmond asked if the differences in study results between rats and mice might be due to the strains that are selected and if there would be fewer differences in rats and mice if there were wider variability. Dr. French said the exact sequence of each strain would have to be known to determine that. Other researchers have explored the same data and found nuances of the mouse genome. There is an advantage to expanding the number of strains to increase the diversity. The HSB has found variation in 18 strains in the benzene ADME studies. There are outliers that may necessitate increasing the number of strains to understand the full range of variability

Dr. Teeguarden asked how the NTP would select strains for testing if there were complete information on susceptibility in all strains of mice, rats, and fruit flies. Dr. French said, at present, industry, academia, and government agencies use a strain of a species and try to extrapolate it to a heterogeneous human population. Few studies have been done on more than a few strains of mice or rats at a time; those studies show quite divergent phenotypes. The relationship between different species shows phylogeny and how genes have evolved over time.

The necessary experiments must be done to define the full range of response and how it might simulate human populations. Dr. Portier asked about the ~ 100 citations of the *Nature* paper based on phase 1 of the Perlegen project. Dr. Johnson said some papers examined particular genes regions for associations and some studied particular types of genetic disorders, e.g., neurological defects. Dr. Kleeberger asked what additional information could be obtained from 50 -100 newly derived inbred mouse strains from the local NC area compared to the data obtained from the Collaborative Cross. Dr. Johnson said the Collaborative Cross is based on eight genotypes, all of which come from laboratory strains that have been in the laboratory environment for many generations. Much of the variation present in the wild is no longer present in members of the eight-way cross; therefore, it is important to isolate some new strains from nature.

Dr. Nagarkatti asked about false positives and negatives in the genome-wide association studies. Dr. Johnson said data could be reanalyzed, adjusting for the false positive and negative rate. False positives have been minimized in the way the analyses were done; the accuracy of the base calls is over 99%. If the data were

reanalyzed, single nucleotide polymorphisms (SNPs) would be gained, but some confidence would be lost in those identified.

Dr. Threadgill commended the NTP on the project and said it has led to many new avenues of research. He asked what is known about the amount of genetic diversity discovered in the resequencing project relative to what is present in human populations. Dr. Johnson said he expects it would model human populations much better. The sampling that has been done on natural mouse populations has been done on individual animals in the wild, not on strains produced from wild animals. Very few wild animals have been examined from very few locations.
Dr. Friedman-Jiménez asked about gene-gene interactions in determining phenotype. Dr. Johnson said they are extremely important and would come out of gene-phenotype associations, once there is the ability to look at the fine detail of the genotype.

Dr. Howard questioned using 50 to 100 strains from NC, rather than getting them from different part of the world and about the human genome project as a parallel to this project. Dr. Johnson said obtaining animals from other parts of the world would be the second step and that most research resources are going toward human sequencing driven by pharmaceutical and insurance companies. No comparable effort is being expended to sequence the mouse genome.

Dr. Birnbaum said over 1000 human genomes are being screened. As the price comes down, potentially costing only $1000/person, more genomes would be sequenced. The data would go into databases, which would be available for data mining. Dr. Howard said databases would be essential to make comparisons.

C. Project 1: Benzene ADME in Genetically Diverse Mouse Strains

Dr. Michael Cunningham, NIEHS/NTP project leader, described the HSB's testing of multiple, genetically diverse, mouse inbred strains to determine the variable range of ADME kinetic parameters using benzene, which has been extensively studied in animals and humans. The hypothesis is that genetic variation inherent in multiple inbred mouse strains would show significant differences in ADME phenotypes that determine bioavailability and toxicity. Preliminary data analyses have established significant differences among the 18 strains, confirming the selection of those strains for in-depth investigation for tissue metabolite distribution. At present, the benzene ADME studies conducted in females of these strains are undergoing final evaluation and quality control for in-depth statistical analysis. If required to improve power for haplotype association mapping (HAM) and statistical analysis, additional strains of inbred mice would be investigated.

BSC Discussion

Dr. Novak, the first lead discussant, said the work was cutting-edge, exciting, and meets the prescribed goals of the HSB. This project has a clear scientific and public health impact. The validity of the approach is compelling and the rationale underlying the project is well defined and strongly supported by preliminary data. He asked about an *a priori* rationale for selecting the strains Dr. French said HSB used the 15 resequenced

strains, the reference strain C57BL/6 (Black 6), and the B63F1 as a reference strain, plus another wild-derived line with no introgression from breeding errors. Those strains had dense genotyping, which had been entered in databases and allowed analysis of genotype/phenotype associations. The selection allowed the use of haplotype associtation mapping (HAM) for the identification of potential quantitative trait loci (QTLs) and selection of candidate genes under the narrow interval peaks to select candidate genes. Additional studies that may be entertained as part of this research project include the comparison of benzene inhalation exposure ADME with that obtained from the oral administration of benzene and modeling of the pharmacokinetic/pharmacodynamic (PK/PD) data. Dr. French said only a small fraction of data has been shown. The 28-day low-dose inhalation study, using ^{12}C and ^{13}C chase, would show potential differences in absorption kinetics, biomarkers (albumin or hemoglobin adducts), and metabolites. Dr. Novak asked whether PK metabolism elimination curves might suggest that the compartment model changes with the animal. Dr. Cunningham said the HSB had been using a one-compartment model, but it may be useful to use a multi-compartment model, given their work on 20 different tissues in 15 different strains.

Dr. Looney, the second lead discussant, asked about specific timelines for various phases of project. Dr. Cunningham said it would take six to eight months to complete this study. Dr. Looney asked what chemical would be used next. Dr. Cunningham said the NTP requests nominations for this type of study. Dr. Looney said adding additional strains do not guarantee an improvement in power. It depends on the variability of the data. Dr. Keith Shockley, NIEHS, said HAM is still in development and power calculations in the context of HAM are scarce in the literature. There are complexities in the studies. Sometimes SNPs are not independent, which complicates the statistics. Approximately 10 strains are necessary to detect a strong genetic effect, but genetic effects are often smaller. Power calculations are dependent upon haplotype structure, which is variable across strains, the genetic effect, the specific statistical algorithm being used, and the significance threshold. The traditional power calculations stem from parametric-type assumptions, which cannot strictly be done in HAM. The HSB has done some calculations making a lot of assumptions. Within that framework, roughly 40 to over 150 strains would give suitable 80% power to detect traits depending on the effect size. HAM requires working with biologists for confirmation and using other types of orthogonal information to improve detection. Dr. Looney requested key publications used in power calculations.

Dr. Eppig confirmed that age-matched animals were used in the studies. Dr. Riviere asked about differences in susceptibility of males and females across strains. Dr. Cunningham said cross-sex comparisons had been done on a few strains. If males are feminized they become refractory to the toxicity of benzene, so the effects are sex hormone-linked. Dr. Birnbaum said, for many chemicals, at puberty significant differences in kinetics develop between genders; in older animals, those differences often lessen. It is not possible to predict which strains would have more variability.

Dr. Teeguarden suggested adding standard error bars in the graphs to get a sense of mean values and precision. Dr. Cunningham said the sample size was five. Dr. Teeguarden said the variability is a factor of 10, greater than expected for PK studies, which indicates the power of these types of studies. He asked how the data could be used to help the NTP select a strain and if the most sensitive strains were used. He asked how data would be used along with genotyping data to identify genes. The data indicate that both absorption and metabolism are involved. Dr. Cunningham said the HSB has a great deal of data on few strains, which may not be the right ones to model all human genetic diversity. Dr. Birnbaum said the "take home lesson" is that the standard practice used by regulatory agencies to split a 10X uncertainty factor between kinetic and dynamic is not going to be protective because of the variability in mice or probably human populations. Other data also argue that point. A 3.3X factor for kinetic differences does not cover intraspecies variability. This is an important message to communicate quickly.

Dr. Threadgill asked about data from human populations and about strains like BALB/c that fall between high and low strains. Dr. Cunningham said there is a great deal of variation in human studies. It would be important to determine what metabolites are in BALB/C mice. BALB/C or C57 could possibly be used as controls or founder strains. Dr. Threadgill said the inbred strains have population substructures that limit the power of detection and increase the false positive rates. It is built into the genetic structure of the populations. He asked if it were possible to use the detailed SNP maps to interrogate the population substructure and sharing of alleles. Dr. French said the HSB has not yet been able to create a database to interrogate the whole data set.

Dr. Friedman-Jiménez asked about using the increase in area under the curve (AUC) to predict leukemogenesis. Dr. Cunningham said the HSB is using those to look for genetic toxicity markers. Dr. Eastmond confirmed that AUC is total radioactivity. Dr. Cunningham said the HSB is using chromatography to isolate different metabolites. Ms. Rudel said it would be good to see a similar kind of distribution for variation due to life stage, route of exposure, or co-exposures. This approach would be a good way to examine mixtures.

D. Project 2: Benzene ADME Phenotype and Haplotype Association Analyses

Dr. French, project leader, explained that data from Project 1 is being used to perform HAM with ADME kinetic parameters (AUC, Cmax, Tmax, or CL_F) and to determine the optimal ADME study design for quantitative genetic analysis. The hypothesis is that individual strain genetic variation would result in significant differences in benzene-induced ADME kinetic parameters in multiple strains of mice that reflect variation and orthologous genes observed in the human population. Preliminary results show significant strain variation. The kinetic parameter data obtained from the males of the 18 sequenced strains are currently being analyzed by HAM to determine if quantitative trait loci can be detected and candidate genetic variants can be identified for functional validation.

BSC Discussion

Dr. Threadgill, the first lead discussant, said the project seems highly appropriate for the HSB as part of its mission to investigate the causes of toxicant response variation. The data set should challenge some paradigms. Regarding the weak correlation between clearance and AUC, there is confounding interaction when they are mapped individually due to two separate genetic drivers. The variance cannot be partitioned out. Dr. French said they might have to repeat some measurements using serial sampling to get more robust inter- and intra-animal estimates of the variation. Dr. Threadgill suggested looking at haplotypes of all the known genes and the variance described by them and how much is left unexplained. Understanding whether variation in ADME phenotypes is due to phase 1 or 2 drug metabolizing enzyme genes or other as yet unidentified genes is important. Dr. French said Dr. Auerbach has proposed 45 candidate genes for that study.

Dr. Threadgill said it was important to match the diets of the mice and humans. Diet may interact with exposures to influence phenotypes, directly or indirectly, in a strain-dependent manner. Hence, the relevance to human metabolic state needs to be considered. Dr. French said there is a mandate to use standardized diets in NTP testing. Diet would have significant effects, e.g., when using the metabolic syndrome diet. Dr. Threadgill said body composition is very different between mice and humans, which make comparisons difficult. Dr. French said the basal expression in white fat and kidney is very different compared to all the other tissues.

Dr. Looney, the second lead discussant, asked about serial sampling of small volumes of blood over time over exposure levels. The idea of “serial sampling of small volumes of blood over time at different exposure levels” as an alternative to discrete sampling of small numbers of animals at individual time points, is certainly well conceived and could potentially provide more “bang for the buck” by eliminating the effect that between-animal variability can have on the statistical comparisons of the time points. However, care must be taken to properly account for the within-animal variation that naturally results when serial sampling is performed (the repeated measures effects). Dr. French said the HSB would work with contract laboratories to develop methodologies. HSB followed the conventional ADME methodology. The HSB can use fewer animals and a greater number of inbred strains if precise quantitative measurements are used over time to study blood kinetics.

Dr. Teeguarden said the presentation was a very believable proof of concept. The NTP has the data, tools, and knowledge base to meet the objectives. He expressed concern about significant confounders such as use of total radioactivity. There are 10-15 different ways to get same AUC for total radioactivity. If the AUC is not correlated with the toxic metabolite in the bone marrow, the answer in the end may relate only to ADME of parent compound and not at all to what is driving the carcinogenicity. It would be better to focus on bone marrow ADME. Using the one compartment model causes the data to be tied together. From a biological perspective, the statistical analysis is not powerful enough to see those relationships or there is confounding by their interdependence. Power can be improved by replacing the one compartment model with one that specifies blood flow to organs, renal elimination, and metabolism. He

suggested carrying out sensitivity analysis to quantitatively determine how blood flow affects AUC. That could be folded into the algorithms that are used to make the connections between the ADME and the physiological or genetic parameters.

Dr. Janzen concurred that the data are impressive and asked about studies of epigenetic variation. Dr. French said those would be incorporated in the future. Dr. Nagarkatti was also enthusiastic about the data and was glad to see the NTP moving on to more cutting edge work. She asked if separate subtypes of cells would be used in the bone marrow cultures. Dr. French said a stromal multi-potential compartment of stem cell type would be created using *in vitro* systems.

E. Project 3: Benzene Low Dose Inhalation Induced Hematotoxicity and Genotoxicity Phenotypes and Haplotype Association Analyses

Dr. French, NIEHS/NTP project leader, explained that data from Project 1 is also being used to design a low-dose benzene inhalation study for quantitative measurement of hematotoxicity and genotoxicity, and to determine the association between benzene oral ADME kinetics and inhalation toxicity. The hypothesis is that genetic variants (SNP and/or structural) between strains determine the exposure-level-dependent, tissue-specific metabolism of benzene and tissue-specific benzene toxicity. To estimate benzene hematopoietic toxicity and circulating blood cells in each exposure group, the HSB would phenotype inbred strains of male mice exposed to low levels of inhaled benzene for 28 days and examine the effect of genetic variation on hematopoietic toxicity and genotoxicity. The C57BL/6, C3H/HeJ, and their F1 hybrids are currently being exposed to validate benzene vapor generation, target exposure levels, and practice collection of tissues for analysis prior to initiation of investigation of 34 inbred mouse strain as a model for the human population.

BSC Discussion

Dr. Novak, lead discussant, said this is a logical extension of the projects discussed earlier and it addresses the goals of the HSB. The project is relevant, the rationale is well developed, the approach is well conceived and justified, and the project aims are logical. It is extremely exciting, valuable data, from both a PK/PD perspective and a genetic susceptibility perspective. He questioned the paradigm of exposures for extended periods of time, with the assumption that this protocol would parallel ADME results from the oral administration. He suggested doing a trial ADME nose-only intermediate dose exposure to see if a parallel result is obtained in a few strains. Dr. French said the HSB considered doing the studies in parallel. There is a window of opportunity in the inhalation facility. The HSB would conduct 28-day studies as proposed and then use those results to design nose-only inhalation ADME studies.

Ms. Rudel said the 100-ppm exposure seemed high and thought the HSB should consider dropping it lower. Dr. French said 100 ppm was used because the HSB has unpublished results suggesting that that exposure is robust for development of myeloid leukemia and T-cell lymphoma. If longer-term inhalation carcinogenicity at lower doses were done in the future, this study would be a good reference. Previous studies at the

Hamner Institute used 100 ppm and there were not enough chambers to add a 5th dose. If effects were seen at 1 ppm, it would corroborate human epidemiology studies. Dr. Novak asked if a 28-day exposure would cause significant hematopoetic toxicity. Dr. French said at steady state levels effects on micronuclei genotoxicity are seen at 14 days in this dose range; the HSB staff wants to go longer because they do not know how much variability there might be among strains.

Dr. Eastmond said one strength of the proposal is that it parallels a number of molecular epidemiology studies of benzene-exposed workers. Comparisons are made of micronucleus frequencies and toxicity of bone marrow, but measurements of plasma benzene metabolites and hemoglobin adducts are also used. The information is from several studies. One-ppm exposure to mice and humans may be equivalent, but due to mice breathing more rapidly, internal dose is likely to be different. It is essential to measure adducts in hemoglobin to make direct comparisons between what is seen in workers and in animals.

Dr. Nagarkatti asked about potential dermal exposure of the mice while in the chamber for 6 hours. Dr. French said dermal exposure would occur during the 6 hours, but the chambers are degassed following exposure to reduce gas levels as quickly as possible. Mice would have some exposure from licking and from residual aerosols.

F. Project 4: Studies to Identify Environmental Cardiotoxins and Susceptibilities to Cardiotoxicity

Dr. June Dunnick, NIEHS/NTP project co-leader, described the toxicology studies to develop sensitive mouse model for the identification of environmental cardiotoxic agents and susceptibility to cardiotoxicity. Components of this program include identification of sensitive and specific biomarkers of cardiotoxicity, identification of more sensitive models for cardiotoxicity testing, and discovery of genetic determinants of chemical-induced cardiotoxicity (CICT). In the next phase, susceptibility to cardiotoxicity in three mouse strains (C3H/HeJ, C57BL/6J, and B6C3F1/J) with divergent SNP patterns in genes critical to heart function would be studied using two mechanistically distinct cardiotoxins, bis(2-chloroethoxy) methane (CEM) and ephedrine/caffeine (E/C). Cardiotoxicity would be quantified using serum biomarkers, ECG endpoints, heart gene transcript patterns, and histopathologic evaluation of heart lesions. In initial studies under control conditions, cardiac phenotypic variation in three strains of male mice (C3H/HeJ, C57BL/6J, and B6C3F1/J) helped identify candidate genes for heritable cardiomyopathy. If warranted, cardiotoxicity studies in up to 34 mouse strains would be considered based on the results of the three mouse strain studies.

BSC Discussion

Dr. Eppig, the first lead discussant, asked about the initial estimates that use of 34 strains would provide 60 to 80% power to detect a genetic association of CICT and the effect of adding or removing strains. Dr. Dunnick said the power calculations were uncertain at this point, but with data from an additional mouse strain, the power would be determined. Dr. French said the 34 strains are a compilation of a mouse phenome database that was used for some of the same measurements. If phenotype were

significant enough, there would be sufficient power to detect effects in this project. Dr. Eppig asked for more detail on the HSB's approaches to filtering gene candidates and suggested using a greater variety of methods to filter and prioritize candidate biomarkers and detect modifier loci. Dr. Auerbach said the HSB tried to leverage a lot of different data originating from HSB and from the UCLA expression Quantitative Trait Locus (EQTL) data now in a gene network database. HSB searched databases for cardiac disease as a literature enrichment analysis. The HSB looked at the gene atlas database for enriched expression of certain genes at a basal level in the heart. It was an iterative process that layered more and more orthogonal data. He said the data are not implying causal relationships.

Dr. Eppig asked if HSB were going to follow-up on the observation that the three candidate genes for cardiomyopathy in C3H/HeJ mice had no differences in the protein coding sequences. Dr. Auerbach said a back transfer is possible, but the HSB would have to justify genes more thoroughly. He said gene trap consortiums have knocked out many of these genes and it may be possible to get access to those strains. The issue of validation is ongoing and would be addressed on a case-by-case basis. He said it is important to look at genes for potential biomarkers in mutant models. The ideal is to have something that is directly related. Diet is an extremely important contributor to cardiovascular disease, so it would be essential to consider modeling the diet in a large-scale study. Dr. Dunnick said the HSB could screen heart tissue from 200 strains at the University of North Carolina (UNC) to look for background cardiac myopathy. Dr. Eppig said the initial project is encouraging and provides a solid proof of concept. The larger project, including additional strains, as outlined, should yield some very interesting cardiomyopathy susceptibility genes and produce useful biomarkers.

Dr. Threadgill, the second lead discussant, suggested looking at the Copenhagen (COP) Project as a way to validate some of the findings. It is important to look at the genes of biomarkers in some of the mutant models; the ideal is to have markers that are directly related (e.g., troponin 1) and not just correlated. Dr. Dunnick said the HSB is considering using the COP strains and screening the 200 other strains from UNC for cardiac myopathy. Dr. Threadgill reiterated that diet is a major contributor to cardiotoxicity. The diet of the strains in a large-scale cardiotoxicity project is very important in order to model the diet of human populations to ascertain the synergies that are occurring between exposures and diet. Dr. Dunnick said the Aging and Disease Phenotyping Project would incorporate different diets to study the effects on aging and cardiac disease. Dr. Auerbach said the C57 strain is susceptible to high fat diets and atherosclerosis. A preexisting cardiovascular condition due to a high fat diet would change the response metric to a cardiotoxin. Dr. Dunnick said even if cardiotoxicity is not measured in background cardiotoxic lesions, a low dose could still make an animal susceptible to a subsequent insult, such as ischemia. Occult toxicity can occur with cardiotoxic agents that are not normally observed. Accumulation of background occult effects may explain sudden cardiac events in humans.

G. Project 5: Aging and Spontaneous Disease Phenotypes in Selected Inbred Strains

Dr. French, NIEHS/NTP project leader, described the HSB efforts to establish a benchmark reference database on aging and disease in 10 genetically diverse mouse inbred strains (129/SvlmJ, A/J, C3H/HeJ, C57BL/6J, CAST/EiJ, NOD.B10H2^{b}/LtJ, NZO/HiLtJ, PWK/PhJ, WSB/EIJ, and B6C3F1/J). HSB investigators hypothesize that the genetic diversity in the 10 strains selected would show a significant age-related range of spontaneous disease and functional phenotypes that would aid selection of strains for toxicology and carcinogenesis research and testing. Aging and survival analysis (spontaneous disease) studies are in progress. Three protocols for identifying and developing biomarkers at specific life stages, and functional analysis of cardiopulmonary functions are in progress and under peer-review.

BSC Discussion

Dr. Looney, lead discussant, thought developing the public benchmark reference database described in this project would certainly aid scientists in selecting the most suitable strains for chemical toxicology and disease studies for extrapolation across species. The rationale for the choice of the 10 inbred strains was clearly spelled out. More detail on the choice of statistical methods and power calculations would be helpful. The discussion of the advantages associated with using different paradigms for comparison of data was somewhat confusing. The statistical power resulting from a two-factor (exposure x strain) design depends on the between and within variance components. If significant interaction between exposure and strain is present, the power of the test to compare the exposure groups would tend to be decreased in the two-factor design since simple effects for exposure (based only on the 10 animals in each strain) would be tested rather than main effects (which would be based on 80 animals per exposure group, ignoring strain). Dr. French said the interaction between exposure and strain is a hotly contested issue. Most confounding problems in doing multiple strain studies are due to fractioning of mice allocated to an exposure group based on strain. It helps minimize type 1 and type 2 errors, but unexpected interactions can occur and it has not been tested previously. It would be problematic to set maximum tolerated doses for each strain because the genetic differences would not be tested. A common metric could be obtained by testing at 75% of the UAC across the strains. Dr. Looney said pilot data are needed to detemine if it is likely that there is a statistically significant interaction between exposure and strain.

Dr. Threadgill said an ideal experiment would be to test diet effects, especially to assess the background characteristics of strain (e.g., cardiovascular disease). It would be important for testing diabetes and obesity. The NTP uses the NTP 2000 diet to minimize kidney effects due to high protein. The NTP did consider using a high fat metabolic syndrome diet. Short-term studies may be warranted.

H. Project 6: Development of a Short Term Cancer Bioassay using Multiple TRP53 Haploinsufficient F1 Inbred Strains

Dr. French, NIEHS/NTP project leader, explained that the database described in Project 5 would be used to prepare for multiple-strain toxicology and carcinogenesis studies and to develop and conduct a short-term cancer bioassay using multiple p53

haploinsufficient F1 inbred strains. The hypothesis is that (1) tumor spectrum, prevalence, and latency, (2) transcript or metabolomic expression profiles, or (3) expression profiles (corroborated by copy number variation) would segregate according to the haplotype of p53 haploinsufficient F1 hybrid isogenic lines selected on the basis of genetic variation in DSB repair genes. Preliminary data in C3B6F1 p53 haploinsufficient mice have shown a non-random allele-specific loss associated with Melm3, Trp53, and Rad51c genes. To date, survival and tumor phenotypes observed by histopathology are significantly different among the four completed F1 hybrid mice studies. Information from these studies would allow the NTP to select inbred strains to perform pre-chronic and chronic multi-strain toxicity and disease studies with NTP nominated chemicals.

BSC Discussion

Dr. Eppig, lead discussant, clarified that both normal and tumor tissue would be assessed and the HSB would be using CGH to detect rearrangements/deletions across the genome in addition to specific changes. Dr. French said CGH is a powerful tool. The HSB would also be using the new mouse diversity array to get allele-specific loss and information about epigenetics. He explained that the studies are for 39-weeks using the NTP design. Initially, studies were done for 26 weeks, due to concern about a high incidence of sporadic tumors in controls. Thirty-nine weeks is not too long to increase sporadic disease in controls and it helps eliminate potential false negatives. Dr. Eppig applauded the F1 design. She asked why the TM1 BRD p53 knockout is being used. Dr. French said the NTP has been using it since 1992, and it is covered under patent.

Dr. Eppig said this is a strong concept with very promising initial data showing a non-random allele-specific loss associated with Melm3, Trp53, and Rad51c genes. She encouraged use of F1 type crosses for looking at genome changes and LOH in tumors. Specific information on benchmark accomplishments, future goals, and timelines would be useful.

I. Project 7: Alkylanilines Class Study

Dr. Scott Auerbach, NIEHS/NTP project leader, described the Alkylanilines Class Study, which has the goal of quantifying the degree to which genetic variation influences the genotoxicity of select alkylanilines and identifying alkylaniline genotoxicity quantitative trait loci. A subset of alkylanilines was nominated to the NTP for toxicological characterization because of the potential for widespread human exposure, the limited availability of published toxicological data for this subclass of alkyl-substituted anilines, and their structural similarities to two known animal carcinogens. The studies have the potential to identify currently unknown genetic loci that modify the carcinogenic risk associated with arylamine exposure. Initial studies to evaluate all 14 alkylanilines for genotoxic potency using an AS52 cell assay are currently being designed.

BSC Discussion

Dr. Novak, lead discussant, acknowledged the merit and validity of the proposed research and the potential scientific and public health impact. The approach is well

conceived, although the proposed studies with primary cultured hepatocytes should consider the variability of the primary cultured hepatocytes and incorporate some measure by which the various cell culture preparations may be deemed viable (e.g., CYP1A2 levels; response to H_20_2 for the Comet assay as well as repair time) for use in examining DNA adduct levels. Widely disparate levels of the enzymes required for alkylaniline metabolic activation and DNA adduct formation may be misleading when associated with the genetics of the strains being examined. Some primary cell preparations are not metabolically viable, even though there is no evidence of aberrant morphology or cell death. The media, levels of insulin and other factors (e.g., time prior to treatment) would be extremely important in this research, as the level of enzyme activity would be affected by cell culture conditions and would reflect the overall functional capacity of the cell and the DNA damage. Since genetics may also govern differences in DNA repair, the Comet assay could be used to assay the time required for cellular DNA repair. Future extensions to exam genetic associations with DNA adducts are appropriate, although hepatic tissue as well as the bladder tissue should be the target of such research.

Dr. Auerbach agreed that the variability in hepatocyte cultures would be a major issue. An alternative approach would use S9 fractions and do metabolic activation with naked DNA; however, this would not assess DNA repair. Pilot studies would be done before the HSB pursues multi-strain analyses.

Dr. Eastmond asked if the prostaglandin A synthase pathway is involved in the activation of these compounds. His concerns were related to the concordance in the mechanism of bioactivation in the liver and bladder. Dr. Auerbach said bioactivation by prostaglandin H synthase, PHS is a possibility and it is likely that adduct formation scales between tissues, which means that a chemical causing relatively high level of adducts in the liver would probably also cause relatively high levels of adducts in the bladder, indicating a similar scale of bioactivation between the tissues. By extension, a similar mechanism of bioactivation likely is occurring in both tissues.

J. Project 8: Impact of Sex and Strain on the Performance of Genomic Signatures of Hepatocarcinogenesis

Dr. Auerbach, NIEHS/NTP project leader, described the development of studies to determine the impact of sex and strain on the performance of genomic signatures for predicting hepatocarcinogenesis. Male F344/N rats, exposed to 30 distinct chemical treatments, would be used to create a 4-class hepatocarcinogenicity prediction model; the 4 classes are genotoxic hepatocarcinogens, non-genotoxic hepatocarcinogens, hepatotoxic non-carcinogens, and non-toxic non-carcinogens. The classification accuracy of the model would be tested on female F344/N rats and on male and female Sprague Dawley (SD) and Wistar Han rats. Chemical and dose selection are underway and initial studies should be started this fiscal year. The HSB would use a similar approach to determine if predictive signatures can accurately classify gene expression changes in the mouse.

BSC Discussion

Dr. Robins, lead discussant, said the research is exciting because it brings a male/female comparison of carcinogen-mediated differential gene expression into consideration when discussing genetic susceptibility. She asked for clarification about what tissue RNA would be isolated, how many genes are differentially affected in the gene expression signatures, and if the signature genes vary genetically among strains. Dr. Auerbach said the left lobe of the liver would be sampled. The minimal signature was between 3 - 50 genes under these conditions, although thousands of genes are different using an ANOVA. He said information could be found in the CEBS database [available at http://www.niehs.nih.gov/research/resources/databases/cebs/]. Dr. Robins said the mouse liver studies are skewed by the androgenic effect and it is not clear how the pronounced sexual dimorphism in rodent liver gene expression is taken into account in this study. Dr. Auerbach indicated that the proposed studies are designed to answer that question. He does not anticipate that the sexual dimorphism would produce profound effects on signature accuracy primarily because the signatures are populated with genes that typically do not exhibit sexual dimorphic expression

Dr. Threadgill asked how the outbred SD and Wistars would affect the gene expression signature. Dr. Auerbach described the portability of signatures across species, specifically with genotoxic chemicals, indicating that the genes predictive in the mouse for carcinogenesis are the same genes that are predictive in rats. Dr. Threadgill explained that he was referring to a cohort with a genetic makeup that is resistant to liver tumors, introducing the possibility that animals selected for study may reflect a genetically susceptible or resistant fraction of the outbred population. This has the potential of creating a disconnect between what the models predict and the outcomes of a bioassay. Dr. Auerbach said a potential solution is to genotype the animals used in a bioassay to assure genetic equivalence to animals used in the predictive genomic studies. The selection of strains is to help the toxicological community, so it is a question of relevance and value to community. The BSC was concerned that the outbred rat strains lacked the genetic homogeneity of mice, which would likely be potential confounder in the derivation and application of genomic signatures. Thus, it is difficult to link gene-expression signatures with terminal cancer phenotypes.

Dr. Portier asked if the training set data collection is at the same time as the collection of validation data. Dr. Auerbach said it would be an iterative process, with all strains done at once. The arrays would all be run together. In the end, the HSB scientists would create the models and test the data. Dr. Portier cautioned about the temptation to look at a whole data set to select what would go into the model. A good design is to have a truly independent training set and a truly independent validation set. Dr. Auerbach said it might be possible to get samples from other researchers that could be used to evaluate the accuracy of the signatures independently. Such samples would be acquired after the creation of the signatures from the proposed studies as to avoid the bias suggested by Dr. Portier.

Dr. Novak asked if signatures had been published for human hepatocellular carcinoma. Dr. Auerbach said two of the genes that are most informative to the predictive models

are commonly deleted in liver cancer and other genes that populate the signatures are found differentially expressed in a wide array of tumors, for example *Mybl2*.

K. Summary and Discussion of the HSB Program

Dr. French briefly summarized the eight HSB research projects, which are in various stages of completion. He mentioned four projects in the planning and review stages: (1) multiple strain toxicology and carcinogenesis pre-chronic and chronic studies, (2) idiosyncratic drug reactions with two agents, (3) creation of a panel of 35 mouse lymphoblastic cell lines (LBCL) for use by the HSB and Biomolecular Screening Branch, and (4) mitochondrial toxicity screening to determine respiratory capacities and toxicity effects across the panels *in vitro*. Dr. French anticipates completion of two to three projects per year and the replacement by similar new projects each year. The HSB would be working with the NTP study design teams to incorporate new tools where warranted. Dr. French said he appreciated the time and effort of the BSC reviewing the HSB program and added that the input was very instructive and valuable to the program.

Dr. Robins said the HSB is highly enthusiastic about the program, which as a whole has made enormous progress in accomplishing its goals. Applying contemporary genetic approaches to address current limitations in toxicological studies is highly relevant and directly applicable to human populations, in which such experiments cannot be conducted. This unique program is poised to take maximal advantage of its activities and provide results to advance human health. Dr. French said the program took a long time to develop. The Perlegen project is now obsolete, but it initiated a lot of effort among many groups; the data have been downloaded into multiple databases and is a valuable tool.

Dr. Bucher thanked the BSC for their efforts and Dr. Threadgill for serving as an *ad hoc* reviewer for the HSB project.

VIII. Nominations and Proposed Research Projects

A. NTP Testing Program: Nominations and Proposed Research Projects

Dr. Scott Masten, NIEHS/NTP, described development of NTP research projects. Toxicological research programs are developed in response to external and NIEHS/NTP nominations. There are multiple levels of review to determine merit and priority for studies. Selection for study is primarily on the basis of production level and uses, known or anticipated human exposure, suspicion of toxicity based on chemical structure or existing data, availability of adequate toxicological data, extent of public concern, and utility of additional studies for public health decision-making. He briefly reviewed the NTP study nomination review process. The NTP staff has prepared draft research concepts for five new nominations with proposed approaches to address toxicological data needs for each substance or issue.

The five draft research concepts for review are: butterbur, evening primrose oil, hydroquinone, silica flour, and valerian. Each research concept outlines key issues,

data gaps, and hypotheses, and/or specific aims that the program plans to address.

The charge to the BSC is to review and comment on draft research concepts and determine whether the proposed research projects are an appropriate use of NTP testing program resources. The BSC is asked to comment on the clarity and validity of the rationale for the proposed research program, the merit of the program relative to the goals of the NTP, the scope of the proposed program and its appropriateness relative to the public health importance of the issue under study, and the priority of the proposed research program.

Dr. Bunton inquired about the volume of nominations received by the NTP and what percentage is brought to the BSC. Dr. Masten said the NTP receives approximately 10 nominations from the general public each year, in addition to substances proposed by NIEHS/NTP staff and other Federal partners. A fairly low percentage of the publicly nominated compounds are brought for review by the BSC because often the requests are for regulatory intervention. Dr. Bucher said there is a small backlog of nominations, but the higher priority substances are all brought before the BSC. Dr. Birnbaum is re-staffing the Washington NIEHS office; one of the new hires will become more integrated with the regulatory agencies to help them formulate nominations that would be directly related to addressing regulatory issues, which should create a stream of nominations. Dr. Bucher said Dr. Paul Howard, NCTR/NTP, has done an outstanding job coordinating nominations from the FDA, a principal group for nominations. Dr. Howard clarified that all of the FDA nominations are presented to the BSC and are fully vetted by the FDA before they are presented to the NTP.

B. Research Concept for Butterbur

a. Presentation

Dr. Howard, NCTR/NTP, presented the research concept for butterbur, which comes from several species of *Petasites*. Butterbur is a dietary supplement/herbal remedy, used for centuries to treat a wide range of conditions including pain, headaches, fever, skin ulcers, urogenital/gastrointestinal diseases, coughs, asthma, allergic rhinitis, gastric ulcers, and inflammation. It is a complex mixture containing sesquiterpenes (including petasin, S-petasin, and furanopetasin), fatty acids, aromatics, phytosterols, and other unknown compounds. The “standardized” formulation has at least 7.5 mg petasin/isopetasin per 50 mg extract. The leaves and rhizomes contain pyrrolizidine alkaloids (PAs), which manufacturers claim to eliminate during extraction.

Butterbur has limited toxicology data; no data are available on subchronic, reproductive/developmental, chronic, or initiation/promotion studies. The LD_{50}, established in Wistar rats, is >2,500 mg/kg (oral) and >1,000 mg/kg (i.p.). Butterbur is mutagenic in TA98 and TA100 *Salmonella* assays. *In vitro* studies with butterbur extracts showed inhibition of histamine and leukotriene-induced contractions in guinea pig trachea strips and inhibition of hexosaminidase release, leukotriene synthesis, and TNF-alpha production in sensitized mast cells. Petasin inhibited lipopolysaccharide-induced PGE2 release and MAPK activation in microglial cells. *In vivo* studies with S-petasin demonstrated modulation of endocrine metabolism in rat testicular cells and

Leydig cells and decreases in heart rate. Both *in vivo* and *in vitro*, butterbur inhibited testosterone release and modulated Ca^{++} channels. Clinical studies are available on butterbur that indicate some effectiveness against allergic rhinitis and treatment of migraines and questionable effectiveness against asthma and allergic skin disease. There are no epidemiological studies on butterbur and it is not recommended for people who are pregnant or nursing, allergic to *Petasites*, have liver disease, or are using anticoagulants, barbiturates, or anti-hyperglycemics.

Butterbur was nominated for toxicology studies by the NIEHS due to its widespread use, toxicity of some of the constituents, and general lack of robust toxicity data for risk assessment. The proposed, tiered, toxicity program would establish a consensus butterbur preparation and then complete *in vitro* screening, subchronic toxicity testing, reproductive/developmental toxicity, and carcinogenicity testing, if warranted. The FDA's position is that any preparation containing PAs may be considered adulterated and therefore may be inappropriate for marketing. The *in vitro* screening would characterize marketed preparations, evaluate the activity/toxicity of those preparations, and allow determination of a consensus preparation. The 28-day repeated dose toxicity studies would use standard toxicity endpoints in rats and mice, focusing on cardio-, neuro-, and hepatotoxicity. The developmental/reproductive toxicity would assess pre- and perinatal oral exposure in rats. The 90-day subchronic toxicity would use rats and mice, exposed orally, and would assess serum hormone levels in rats in addition to standard toxicity endpoints (again focusing on cardio-, neuro-, hepatotoxicity). These studies would provide toxicological data to enable the NTP to (1) quantify the toxicity of butterbur and constituents and (2) generate data for useful for FDA to develop a risk assessment of butterbur dietary supplements and herbal preparations.

b. **Public Comments**

Mr. Joseph Manuppello, People for the Ethical Treatment of Animals (PETA), said butterbur, evening primrose oil (EPO), and valerian root extracts are all herbal dietary supplements with long histories of safe use. In each case, clinical trials have already been conducted with these substances and have provided evidence for their safety. The NTP has overstated its concern for their potential toxicities. There is clearly sufficient confidence in the safety of each of these substances to conclude that any remaining concerns for their potential toxicities can be appropriately investigated through clinical trials and post-marketing surveillance studies. The unnecessary and irrelevant animal testing programs proposed by the NTP would waste time and resources in addition to animals' lives.

Mr. Manuppello said a chronic toxicity study in rats for butterbur exists and he questioned why FDA considers the study incomplete. It is an unpublished 26-week toxicity study in rats from 2001 by the German company Weber & Weber, manufacturers of the commercial butterbur preparation Petadolex®. This study is described by Danesch and Rittinghausen (2003), employees of Weber & Weber, as a chronic toxicity study performed in 200 Wistar rats for a period of 26 weeks. The study was conducted in accordance with ruling European Economic Community Directives, and International Conference on Harmonization and Organisation for Economic

Cooperation and Development Guidelines. Principles of Good Laboratory Practice (GLP), as specified by international law, were strictly followed. The study established a no observed adverse effect level (NOAEL) well above and at an adequate safety margin from the recommended dose in humans. PETA contacted Dr. Danesch, but he was unable to make the results of the study available, citing confidential business information. Mr. Manuppello said the NTP would also be unable to make this study available to the BSC or public on the NTP's website. Mr. Manuppello said it is unacceptable that such highly relevant data would be ignored over this procedural challenge. The NTP must find a way to obtain and thoroughly review the results of this study before initiating clearly repetitive animal tests.

Butterbur has a long history of safe use. It's been used since the 17th century to treat a variety of medical conditions. A post-marketing surveillance study of 580 patients treated for allergic rhinitis was conducted with a commercially available butterbur extract in Switzerland. Käufeler *et al.* (2006) found the overall incidence of adverse events to be very low and similar to the occurrence of adverse events in the placebo group, consisting primarily of mild gastrointestinal symptoms. The authors concluded that butterbur extract, confirmed by three Good Clinical Practice trials and two post-marketing surveillance trials, to be safe and effective. Similarly, Guo *et al.* (2007) conducted a systematic review of five databases to evaluate the quality of the available studies on butterbur and other herbal medicines for the treatment of allergic rhinitis. They concluded that most trials found good tolerability and similar adverse events compared with the placebo.

Mr. Manuppello said the NTP's research concept document cited evidence from one clinical case report of fetal death as the result of maternal consumption of butterbur to support its concern for adverse effects in pregnant or nursing mothers. It is important to note that the mother consumed large amounts of herbal tea, which contained PAs, during pregnancy. As stated in the review document, PA metabolites are known to be hepatotoxic. Current processing methods using CO_2 or propane yield butterbur extracts that contain PAs below the detection limit and most commercial products state that the preparations are PA-free or contain no detectable PAs. This known source of potential toxicity can therefore be effectively controlled by analytical methods and appropriate labeling of available butterbur products. Additional animals tests on butterbur are unnecessary for regulating PA contaminants.

Butterbur's long history of safe use and the lack of serious adverse effects reported in safety, clinical, and efficacy trials indicate a low level of concern for the potential toxicity of butterbur products prepared by current extraction methods. As with the other nominated herbal dietary supplements, butterbur's potential toxicity can be appropriately investigated in human studies, which would yield more relevant information.

c. BSC Discussion

Dr. Eastmond, first lead discussant, said the rationale was adequately presented and valid. The proposed research addresses the NTP mission and goal of providing information on potentially hazardous substances to all stakeholders. This research

project could potentially be very significant. Dr. Eastmond rated butterbur's significance and public health impact as medium to high. It is somewhat difficult to assess as so little information is available on the potential hazards of butterbur. It clearly has pharmacological properties and could easily have significant toxicological properties. He concurred with using a tiered approach and recommended conducting initial *in vitro* and *in vivo* screening studies and then making decisions about whether the more comprehensive and expensive chronic studies should be conducted. He said the PAs in the butterbur could have a significant impact on the studies. The post-market studies would be advisable if there were a long history of chronic use of butterbur, but they would be difficult. Epidemiology studies may be useful to identify short-term clinical outcomes, but long-term chronic effects would be difficult to characterize. Dr. Eastmond considered the proposed testing warranted and encouraged use of the tiered approach.

Mr. Janzen, second lead discussant, concurred with Dr. Eastmond. He considered a major outcome to be the extraction methodologies. The proposal was well thought out and he supported the partnership with FDA for the study. He was unclear about the gating events in the tiered approach, what data would drive the decisions and what data would be used to decide on carcinogenicity studies. Mr. Janzen said the project fits with the NTP's mission and he rated it as having a moderate impact for public health. The scope is appropriate and the study uses genotoxicity studies to reduce the use of animals.

Dr. Howard said companies are not required to make their data available. The 26-week Wistar rat butterbur study Mr. Manuppello described is not a chronic study, so it is hard to evaluate. Butterbur has been used historically by pregnant women, but the referenced study did not use perinatal exposures. In many of the studies of dietary supplements, there are no correlative exposure groups in animals. Dr. Howard said the NTP butterbur study would not use a preparation that contains PAs, because the goal is to study the other components of butterbur to better understand human risk. He said every six months the FDA and the NTP jointly evaluate dietary supplement data and decide how to proceed in tiered studies. The public health agencies decide at what point in the tier they have adequate information to decide on a regulatory action, i.e., the regulatory action determines how the tiered approach proceeds.

Dr. Faustman asked for clarification about the use of butterbur by pregnant women. Dr. Howard said no clinical trials are underway to address the reproductive safety of butterbur. There has been historical use of butterbur for dysmenorrhea and currently it is used worldwide due to its efficacy in reproductive age women. Dr. Faustman suggested the NTP provide guidelines for use of the tiered approach in studies. Dr. Howard said sometimes 90-day studies are required to ascertain the need for carcinogenicity studies, whereas for other substances, more short-term studies identify toxicity early, so there is no need for carcinogenicity studies. Dr. Bucher said the NTP would welcome the BSC’s suggestions about what data would be adequate to justify going on to a further study. Dr. Portier said the NTP has points at which the data are reviewed and decisions are made regarding studies. Dr. Bunton asked about the consensus formulation for butterbur. Dr. Howard said it would depend on whether PAs

are present in the marketed butterbur, but he thought the testing would be without PAs. Dr. Portier asked about the cultivation of *Petasites.* Dr. Howard said there is great variability in some herbals due to season of harvest, growth conditions, latitude of cultivation, weather, harvesting methods, etc. He doubted that the test article would be consistent throughout its period of use, thus rendering retroactive epidemiology studies less useful. He said the proposed study would focus on modern usage. Ms. Rudel asked about databases for over-the-counter medications. Dr. Howard clarified that butterbur is not considered an over-the-counter medication, which have different regulatory requirements. There are scattered databases with information on efficacy of supplements, but there is not a consistent toxicology database. A multi-agency program is being considered to collect information about dietary supplements. A new dietary supplement is required to have some information about toxicity, but it is not required for preexisting supplements. Dr. Teeguarden asked about the clinical data on long-term human use of butterbur. Dr. Howard said the studies were constructed to look at efficacy, not hazard identification, so there is a limit to the power of the studies to predict toxicity. Mr. Janzen asked if a regulatory decision could be based on *in vitro* data. Dr. Howard said *in vitro* studies are not sufficient.

C. Research Concept for Evening Primrose Oil (EPO)

a. Presentation

Dr. Dori Germolec, NIEHS/NTP, presented the research concept for EPO, a biennial weed that is a minor oilseed crop used to produce the dietary supplement. It is a source of essential fatty acids, particularly linoleic acid (LA) and γ-linolenic acid (GLA). EPO was nominated by the NIEHS for toxicological characterization based on widespread use in dietary supplements, lack of adequate toxicological data, and concern regarding potential adverse health effects. Human exposure primarily occurs through the consumption of EPO-containing dietary supplements, which are routinely in the top 20 of herbal sales. EPO is licensed for the treatment of mastalgia, premenstrual syndrome (PMS), and prostatitis in the United Kingdom (UK) and is suggested use for treatment of a variety of other inflammatory conditions including atopic eczema, psoriasis, multiple sclerosis, cancer, coronary heart disease, diabetic neuropathy, autoimmune conditions, and gastrointestinal symptoms. EPO is also used during pregnancy for prevention of pre-eclampsia and early delivery, and shortening and stimulating labor. Most human studies have examined efficacy for anti-carcinogenic effects, treatment of neurologic disorders, and modulation of labor and delivery. EPO is well tolerated by most people, but long-term safety has not been evaluated in human studies. Mild side effects include gastrointestinal upset and headache. Some cancer patients have objective improvement following EPO treatment, but results are variable. Use of EPO in nulliparous women was associated with increases in labor time and active phase labor abnormalities. Early suggestions of association with seizures were not supported by review of the studies.

One long-term study has been conducted demonstrating no significant differences in body weight, weight gain, and food consumption in mice, rats, and dogs fed EPO. There were minimal effects in a chronic toxicity study in Sprague Dawley (SD) rats and beagle dogs. A carcinogenicity study by the same group showed no significant

differences in tumor incidence, but a trend for fewer tumors in male rats and all mice. Reproduction studies of EPO exposure in SD rats showed testicular shrinkage or softening; in Wistar rats there were no differences in parturition, birth weight, postnatal growth rate, maternal weight during pregnancy, and fetal or placenta prostaglandin E2 levels as compared to control animals. In Imprinting Control Region (ICR) mice there were increases in body weight, testis weight, and testosterone levels with EPO exposure. There is no evidence that EPO is genotoxic. A reduced incidence of micronuclei when co-administered with benzo(a)pyrene suggest that EPO may prevent DNA adducts. Immunotoxicity studies suggest potential hypersensitivity following exposure to EPO, and there is some evidence of immunosuppression in rodents. The metabolism of essential fatty acids is well understood in humans and laboratory animals. Animal tissues are more active in conversion of LA to longer-chain polyunsaturated fatty acids (PUFAs) than humans. There are limited data on the bioavailability of LA and GLA following EPO treatment and this is part of the proposed research program. A key question is which product or formulation to study. Commercial preparations vary in their fatty acid content and many of them are also supplemented with other PUFAs (such as fish oil) and/or antioxidants. The available toxicology data are extremely limited, but there is little evidence of systemic toxicity following EPO consumption in humans or experimental animals. The pattern of use and reported effects on reproductive endpoints suggest that additional studies in this area would be warranted. There is evidence of immune effects, most being consistent with what has been reported in the literature with regard to essential fatty acid supplementation

The overall goal of these studies is to characterize the subchronic and reproductive toxicity of EPO following oral exposure in rats and mice. A tiered approach with defined specific aims would address this goal. Tier 1 would have two specific aims: (1) conduct studies on the bioavailability of EPO and (2) conduct prechronic toxicity studies in the SD rat and B6C3F1 mouse. The rat studies would be conducted using the perinatal exposure paradigm and would provide information on the potential targets of EPO toxicity. These studies would provide critical information for dose setting in the reproductive and continuous breeding (RACB) study. Tier 2 would follow with additional specific aims: (3) conduct a guideline reproductive toxicity study to examine the effects of EPO on fertility and fecundity in SD rats and (4) conduct a standard immunotoxicity screening panel to assess the potential immunomodulatory effects of EPO. Tier 3 would evaluate the need for additional toxicity studies based on the results of Specific Aims 1 and 2. Information on the number of individuals who take EPO is not readily available. The reported effects on pregnancy and reproductive endpoints may be a cause for concern. These studies would address data gaps in this area. Additional studies would address the lack of toxicity data for EPO and provide much-needed information on its safety for the FDA and the public.

b. Public Comments

Mr. Manuppello, PETA, said the NTP's concern for the potential toxicity of EPO appears to be greater than the evidence warrants. The entire evening primrose plant is edible. Native American tribes consumed it as a staple food and used it medicinally. EPO

consists mainly of essential fatty acids including LA, GLA, oleic acid, and palmitic acid. They are normal intermediates in human metabolism and very unlikely to be inherently toxic. In a 1992 review, Horrobin reported that of 4000 patients involved in clinical trials of EPO for three months or more at doses of 3-6 g/day not a single adverse event occurred significantly more frequently in patients receiving EPO than in those receiving placebo. The studies reviewed included one in which 14 children with cystic fibrosis received doses of up to 20 g/day for one year. Horrobin also noted that about half a million prescriptions for EPO have been dispensed by the UK National Health Service for the treatment of atopic eczema or breast pain and that the incidence of reported adverse events has been far lower than with most drugs with no pattern suggesting that any adverse events were related to its use. Long-term studies in four species of animals in which reproductive toxicity, teratogenicity, and carcinogenicity were evaluated found no toxic effects attributable to EPO. In the research concept, the NTP states that the reported effects on reproductive endpoints suggest that additional studies in this area would be warranted. The NTP claims that labor lasted longer in women taking EPO than in those who did not, citing a retrospective study by Dove and Johnson (1999). These authors state that they could not conduct a valid analysis of variance to determine whether differences in length of labor existed because they observed a wider variation in length of labor in the EPO group. The NTP also claims that the use of EPO was associated with increases in active phase labor abnormalities including incidence of cesarean delivery and vacuum extraction. However, the study's authors noted no effect on cesarean delivery in their discussion and expressed uncertainty over the significance of a marginal increase in vacuum extractor use. They also note that their study was limited by its retrospective design that used client records to assess outcome variables. There was no strict procedure for defining onset of labor or latent phase and no attempt was made to control for other variables, including the use of homeopathic remedies or castor oil, which might independently influence cervical ripening or labor onset. The authors identified a need for replication of their study with a larger, more controlled sample.

Mr. Manuppello said with regard to reproductive toxicity in animals, the NTP cited Leaver *et al.*, (1986) who showed that Wistar rats fed a diet supplemented with a commercial EPO preparation from three weeks of age until mating showed no differences in parturition, birth weight, postnatal growth rate, maternal weight during pregnancy, and fetal or placenta prostaglandin E2 levels as compared to control animals. Effects on male reproductive function in ICR mice orally administered EPO, reported in an abstract, included increases in the number of complete penile insertions during a three-hour period. These effects were apparently not regarded as adverse effects by the study's authors who titled their work, "Improving effects of evening primrose oil on the sexual functions of male mice." What the NTP describes as the demonstrated evidence in the literature of effects on reproductive endpoints in humans and experimental animals does not support the proposed reproductive toxicology study.

The NTP proposes an immunotoxicity screening study to "clarify" the immune functions and cell populations modulated by EPO. As noted previously, EPO consists of PUFAs, some of which have been shown to have immunomodulatory effects. Existing and

ongoing investigations of the effects of these individual constituents are applicable to determining the potential immunotoxicity of EPO as a whole. As with the other nominated herbal dietary supplements, PETA believes that any further investigation should be assigned a low priority and preference should be given to studies of effects in humans.

c. BSC Discussion

Dr. Nagarkatti, first lead discussant, said the rationale was well articulated and clear. The proposed studies are highly significant because EPO has been used extensively for a wide variety of illnesses and also in cosmetics. Performing a systematic toxicological evaluation, in addition to determining immunotoxicity and reproductive toxicity of EPO, is important, as it would shed light on the potential hazardous effects on human heath. There has been a great deal of controversy on the use of EPO. It is grown and used worldwide, so it is potentially a global health problem. The entire plant is edible and composed of a large number of essential fatty acids. The procedure for extraction may confound some of the toxic effects. Since the FDA does not currently regulate EPO, the studies would provide information on potential future regulatory actions. The public health impact of the studies could be high. The initial studies would focus on genotoxicity, immunotoxicity, and ADME; later studies would focus on other toxicities. Inclusion of other routes of exposure, such as dermal, may be beneficial. Additionally, EPO may have toxic effects when administered as a mixture with other compounds.

Dr. Bunton, second lead discussant, applauded the NTP for addressing herbals and dietary supplements. She cautioned that they are not inadvertent exposures to substances, but rather are compounds that have been deliberately taken by humans, often for many years. This should be kept in mind when studying compounds that have been assessed in human clinical trials, which test for efficacy, but also report toxicity. The NTP should not conduct unnecessary animal studies that would not provide additional information beyond what is known from the human clinical trials. She agreed that the proposal was well written and clear. Very limited toxicity was demonstrated in previous animal studies; those data cannot be dismissed. Given the past clinical trial and animals studies in the literature, the value of additional studies as proposed would be limited and may not be a justifiable use of animals or resources. She questioned whether teratogenicity studies had been conducted on EPO. An assessment of teratogenic potential is one area that appears to have limited available data and might be targeted in additional studies. The effects of EPO on the immune system are nebulous; they appear to be immunomodulatory rather than immunotoxic. The potential immunomodulatory effects of PUFA have also been evaluated through human studies and are contradictory; therefore, it is not clear how an immunotoxicity screening program would be of value. The overall significance and public health impact of this research program is low. Dr. Germolec responded that developmental teratogenic effects were not reported in the EPO study. A critical issue with PUFA dietary supplements is the balance between omega-3 and omega-6 fatty acids; alterations in that balance can result in adverse effects in endpoints such as reproduction. The stimulatory effects of EPO are a cause for concern, because unintended stimulation of the immune system is considered an adverse effect that can lead to autoimmune

disease or enhancement of allergies.

Dr. Riviere commented on the mechanism of action that leads to atopic dermatitis. LA is a major constituent of intercellular lipids. It is not certain whether a barrier disruption leads to atopic dermatitis or the converse, but LA preparations effectively treat the dermatitis. Part of the immune action is confounded due to the barrier aspect. LA supplementation is used in veterinary medicine for dermatologic conditions. The method of extraction of EPO (cold extract versus hexane) should be studied to determine if it affects the nature of the preparations and for effects on reproduction.

Dr. Faustman expressed concern regarding the high doses used in some of the previous studies. She supported reproductive testing of EPO and mentioned neurodevelopmental processes that are modulated by fatty acids. Dr. Teeguarden agreed that reproductive endpoints are very important and asked for responses to the public comments. Dr. Germolec said the issue was relying on a single, possibly underpowered epidemiology study. There are ethical issues with conducting additional epidemiology studies of EPO use. The NTP is concerned that EPO has demonstrated effects on reproductive parameters in women and is being taken by a large number of women of childbearing age. Dr. Nagarkatti said in addition to the reproductive effects, the immune system effects should be addressed due to the contradictory data. Ms. Rudel said a limitation of the epidemiology study is that it is retrospective and has uncontrolled confounding. The women who chose to take EPO at the end of their pregnancy are probably different in many ways from the women who did not take EPO. Dr. Bunton said women who are pregnant are not encouraged to take EPO, but are rather discouraged. A rat study has shown no effects on parturition, but the rat study may not be reflective of the human study.

Dr. Portier said he views EPO as a product composed of natural oils that are well characterized. He questioned why a mixture study was not proposed. Dr. Howard said the FDA appreciates any data that indicate toxicity, but also data that indicate lack of toxicity, which is important from a public heath standpoint. There are many dietary supplements on the market, but there are even more supplements that are a mixture of supplements, i.e., complex mixtures of mixtures. One individual can consume many dietary supplements. Understanding underlying mechanisms of action allows an understanding of effects when supplements are used together.

Mr. Manuppello clarified that he was attempting to raise concern with the single human epidemiology study because it was the main source of the NTP's concern.

D. Research Concept for Hydroquinone (HQ)

a. Presentation

Dr. Michael Sanders, NIEHS/NTP, presented the research concept for HQ, which was nominated by the FDA for toxicological evaluation because data are needed to evaluate the risk of exposure to HQ in topically applied consumer products. Specific studies requested are to assess reproductive toxicity, dermal toxicity, and carcinogenicity in animals. HQ and its glucose conjugate arbutin occur naturally in plants and foods. It is

used in industry and in topical medicines to treat disorders of skin pigmentation and for skin bleaching. HQ's mode of action is inhibition of melanin production and selective damage to melanocytes. HQ has been shown be a clastogen and neurotoxic at high doses. There is evidence of reproductive toxicity and carcinogenicity in rodents and dermal toxicity in humans using topical preparations.

The NTP studied HQ in the early 1980s and found no clinical signs of toxicity in dermal 14-day studies; decreased body weights, tremors or convulsions, and mortality in oral 14-day studies; mortality, central nervous system effects, nephrotoxicity, and forestomach toxicity in oral 90-day studies; and increased incidence of kidney adenomas, leukemia, and liver adenomas or carcinomas in oral 2-year studies.

Mechanisms of action of HQ include enzyme inhibition of melanogenesis, inhibition of RNA or DNA synthesis, oxidative stress, and binding to macromolecules. Key issues include dermal exposure and toxicity; the effects of long-term dermal exposure on systemic toxicity and toxicity at the site of application are uncertain. The existing reproductive toxicity data in animal models are conflicting and incomplete.

Specific aims in the proposed study are to: (1) conduct both oral and dermal ADME studies of HQ in rodents, (2) conduct reproductive toxicity studies of HQ in rodents using the NTP RACB protocol with oral exposures, and (3) conduct dermal toxicity and carcinogenicity studies of HQ in rodents.

The significance of the proposed research program would be to provide toxicological data needed by the FDA to evaluate the risk of exposure to HQ in topically applied consumer products. Dr. Sanders acknowledged FDA's involvement in the development of the concept and in future study design.

b. BSC Discussion

Dr. Riviere, first lead discussant, said the rationale is clearly presented. There are conflicting data on numerous aspects of the toxicity of HQ that has prevented an accurate characterization of its hazard, despite its human exposure through various sources including topical preparations for skin lightening. Laboratory animal data, to date, have shown conflicting results, possibly due to strain differences. In some cases, such data were collected after oral administration, which may not be relevant to dermal administration. Recent studies have been conducted, but they have not been published. He deemed the project consistent with the NTP's mission and considered it to have a high priority. Two major issues that need to be address are HQ's comparative metabolism and its potential toxicity after dermal administration. Attention should be paid to potential formulation issues so results are comparable to human exposures.

Dr. Eastmond, second lead discussant, said the rationale for the proposed research program was adequately presented and agreed that it is consistent with the NTP's mission. The comparative studies on absorption and disposition would also contribute to the knowledge base for this type of compound. He rated the overall significance of the proposed research to be moderate to high. There are large numbers of people who

use HQ-based skin lighteners, often for prolonged periods. He mentioned a group of people in Nigeria who regularly use skin lighteners. Generating data on HQ's chronic, carcinogenic, and reproductive effects from dermal exposure would be useful for those making safely decisions about this chemical. He said a 2-year cancer bioassay was previously completed by the NTP using the drinking water route of exposure; however, there are likely important differences in absorption and metabolism between drinking water and dermal exposures, which would justify this study. Route of administration plays an important role in the toxicity of this type of agent and consequently should be an important component of the study. With oral exposure, there is efficient conjugation in the intestine and liver, whereas with dermal exposure there is direct access to systemic exposure and a potential for direct toxicity. He considered the scope of the work appropriate and recommended a tiered approach. The information derived from the ADME studies should be used in the design of the chronic and reproductive studies. The skin is an effective barrier, so it may be possible to estimate what plasma concentrations of HQ would be, making some of the previous studies more useful. He pointed out some inconsistencies in the background nomination document that should be revised. He suggested getting more information about the current skin lightening mixture carcinogenesis study.

Dr. Sanders said the methodologies for the dose formulations would be done with the ADME studies. The NTP may also examine absorption using other vehicles. Dr. Riviere agreed that that would be important. Dr. Sanders said the tiered approach would be used; the ADME studies would be done first and those data would guide later decision points. Inconsistencies in the nomination document were not brought into the concept document. Dr. Eastmond said oral exposures to HQ do not produce DNA adducts, but adducts are formed with metabolites from benzene. Dr. Sanders said radiolabels would be used in the ADME studies to identify DNA adducts in the skin. He said the NTP would make sure not to duplicate HQ studies being conducted elsewhere.

Dr. Carney asked for clarification and justification of the reproductive studies. Dr. Sanders explained that the reproductive studies would be done in rats only. The NTP reproductive studies are very comprehensive and allow flexibility for study design. The earlier studies done by industry were much less comprehensive. Dr. Bucher said the reproductive studies done by the NTP vary depending on the needs of the chemical. Dr. Portier questioned the justification for an oral reproductive study. Dr. Foster said a dermal reproductive study is nearly impossible to conduct and RACB studies are used as the default. Dr. Howard said the study design would be determined by the ADME studies. If FDA receives any additional data on HQ, it would be shared with the NTP. The FDA has maintained a position for many years that dermal carcinogenicity studies are needed to better understand the toxicity profile of HQ. It is the FDA's opinion that none of the information conveyed today changes FDA's position on the need for dermal carcinogenicity data, and the FDA fully supports the proposed studies.

Dr. Faustman asked if the additional data being discussed were from a human pregnancy study. Dr. Masten clarified that the human study alluded to was for an approved drug product containing HQ, a corticosteroid, and retinoic acid.

E. Research Concept for Silica Flour

a. Presentation

Dr. Cunny presented the research concept for silica flour, which is finely ground quartz crystals 1-100 µm in size. It was nominated by a private individual for testing via oral and dermal exposure with an emphasis on immunotoxicity testing. This nomination was based on studies showing that occupational exposure to silica has been linked to higher rates of autoimmune diseases such as lupus, rheumatoid arthritis, and scleroderma. The nominator suggested human oral and dermal exposure might be via cosmetic products and filler in vitamins and pain relievers. There is a paucity of oral and dermal toxicity data on silica in general. Most data on silica are based on respiratory exposure. Occupational exposure to respirable crystalline silica is known to cause silicosis and to be carcinogenic.

The term “silica” refers to the chemical compound silicon dioxide, which occurs in crystalline or non-crystalline (amorphous) forms. Crystalline silica may be found in multiple forms including alpha quartz, cristobalite, or tridymite. In nature, alpha quartz is the most common form and is abundant in rocks and soil. Silica flour is a very pure grade of finely ground alpha quartz. Silica flour is used as an abrasive cleaner and inert filler. It is used in glass, ceramic and paint production, and in skin care products such as exfoliants and wart removers. The extent to which silica flour might be found in vitamins or pain relievers is difficult to ascertain. The ingredient “silica” is listed as an inert in a variety of vitamins and pain relievers, and this generalized term does not provide information on the crystalline or non-crystalline nature of the ingredient. More investigation would be needed to determine if silica flour might be in products such as these.

Occupational exposure to respirable crystalline silica is a known lung carcinogen. Numerous occupational epidemiology studies reported moderately elevated risk for extrapulmonary cancers including in the gastrointestinal tract. Inhalation of respirable crystalline silica is associated with chronic silicosis, usually a nodular pulmonary fibrosis. Other silica-related diseases include pulmonary tuberculosis, chronic obstructive pulmonary disease, chronic renal disease, hyperthyroidism, scleroderma, rheumatoid arthritis, and lupus. There are limited animal data on non-inhalation exposures to crystalline silica; however, some reported effects are focal tubulo-interstitial nephritis following exposure via drinking water (for 4 months) in guinea pigs; adjuvant effect on antibody production following intravenous exposure, non-specific focal hemorrhage in liver and heart, fatty liver, and increased ALT in mice following oral exposure.

Among the key issues for silica flour, it was pointed out that there is a general lack of oral or dermal toxicity data and ADME studies via non-respiratory routes. The NTP has also identified a need for a better understanding of particle size, shape, and other physical/chemical characteristics in consumer products.

The proposed approach would include an attempt to better define potential exposure to silica flour and define particle characteristics in various consumer products containing crystalline silica to choose appropriate test material. ADME studies would be conducted to determine bioavailability, if feasible. Studies with oral exposure would be conducted before dermal exposure, using particles of two or three size ranges/shapes, if warranted. The NTP would attempt to determine if there is a particle size/shape factor in bioavailability. The NTP would conduct immunotoxicity assays using the NTP tiered testing panel and then select a clinically appropriate autoimmune animal model based on the observed effects. General toxicity testing would be considered based on the study's findings.

The NTP studies on silica flour would investigate the potential for oral exposure to crystalline silica to cause non-pulmonary adverse effects including autoimmune disease and would also provide ADME data. The NTP research program would address the data gap for the potential toxicity of non-respiratory exposure to crystalline silica.

Dr. Teeguarden requested confirmation that silica flour is used in pills and cosmetics. Dr. Cunny said there was anecdotal evidence of exposure to silica flour through pills. Dr. Friedman-Jiménez described a recent industrial hygiene survey of a vitamin and supplement manufacturing facility in which silica flour was used as a filler. He encouraged additional inquiry into the use of silica flour as an ingredient in pills. Dr. Portier said the personal care products industry states on its webpage that they use amorphous non-crystalline silicon dioxide in cosmetic products.

b. BSC Discussion

Dr. Friedman-Jiménez, first lead discussant, said the rationale for focusing on immunotoxicology should include more references to epidemiology studies reporting autoimmune disorders in workers with silicosis exposed to respirable silica dust. The rationale for focusing on oral and dermal routes of exposure and not including the inhalation route was not stated clearly enough. In particular, there is a lack of compelling evidence for potential toxicity of silica via the dermal route of exposure. Dr. Friedman-Jiménez provided some examples of occupational studies and *in vitro* studies that should be included in the background material. He suggested a more thorough review of the epidemiologic literature on silica, silicosis, and a range of immune and autoimmune disorders including rheumatoid arthritis, scleroderma, lupus, tuberculosilicosis, and other less common disorders. He said the proposed research program would provide information on potentially hazardous routes of exposure to silica, which would strengthen the science base in toxicology. It is not clear whether this research would involve development and validation of new or improved testing methods. He rated the overall significance and public health impact of the proposed research program on silica flour as moderate for the oral route and low for the dermal route. He recommended adding mechanistic studies of immunotoxicity related to silica flour exposure by the inhalation route.

The main public health impact of autoimmune disorders due to silica is most likely what the epidemiology data suggest, largely among people with pulmonary fibrosis due to

inhaled silica. Mechanistically, the fibrogenic potential of silica may underlie the carcinogenic effects of silica, autoimmune connective tissue disorders such as rheumatoid arthritis among people with pulmonary fibrosis due to silica exposure, and other effects on the immune system such as increased susceptibility to tuberculosis among people with silicosis. He considered the public health importance of immune toxicity of silica flour to be due predominantly to adverse health effects among people with inhalation exposures to silica flour, much more than among the larger number of people with exposure to silica flour by the oral route. Exposure to respirable silica dust and silicosis are still current issues. Silicosis is a debilitating disease and currently there is no treatment. Dr. Friedman-Jiménez provided some statistics on silica exposures and mortality in both the United States and other countries and discussed his experience with silicosis patients. He thought there would be substantial overlap of mechanisms between oral and inhalation exposures and recommended that the NTP add comprehensive testing of silica flour by the inhalation route to the proposed research program. He would then rank the overall significance and public health impact of the inhalation studies, with emphasis on *in vivo* and *in vitro* mechanistic studies, as high. These studies could also facilitate development of chemoprevention strategies and treatments for immune and renal disorders.

Dr. Teeguarden, second lead discussant, said silica *per se* does not exist; silica comes in many forms that are very distinct entities toxicologically. He recommended being more specific in using the word "silica" in the documents, as amorphous and other crystalline forms of silica have very different properties. He said the rationale is clearly written, but he did not find strong evidence for oral exposure or dermal exposure, other than occupational. The presumed hazard is development of autoimmune disorders, but the “evidence” of autoimmune disorders presented in support of the nomination was cited as no more than a “possible link.” Other evidence supports inflammatory effects of silica after inhalation exposure and *in vitro* exposures. Because it is known that systemically absorbed particles are taken up by macrophages and are capable of disrupting function of these cells and causing inflammatory effects, there is mechanistic support for the possibility of other immunological effects. Inhalation exposure typically leads to some oral exposure, so studies of inhalation exposure also include oral exposure. Given that inhalation exposure is expected to result in systemic exposure, the absence of stronger links between inhalation exposure and autoimmune disorders in humans limits the validity of the concern regarding dermal and oral exposures.

Systemic exposure following dermal exposure should be limited as even studies with very small particles show that the dermis is a good barrier to particles. Greater exposure due to oral exposures seems likely, and there are some rodent data that support this conclusion. He questioned the statement in the background materials that amorphous silica is cleared from the body more rapidly than crystalline silica because it is more water-soluble. Clearance is related to the effectiveness of macrophage clearance, dissolution, and particle size. Dr. Teeguarden gave weak support for the nomination and said the proposed research program has merit under goal 1 because it would provide information on potentially hazardous materials and strengthen the science base in toxicology. He rated the significance as limited, but said it could be

strengthened if the document were included a systematic review of the available information on oral exposure.

Dr. Teeguarden considered the scope reasonable and said the exposure and data on the physical/chemical characterization of materials would be crucial for defining what to study and at what exposures. He concurred that bioavailability studies should proceed toxicology studies. It would be essential to determine if there are particle size dependencies in bioavailability, distribution and/or response. He suggested the NTP attempt to compare the amount of material absorbed by the oral and dermal routes to what goes into the systemic circulation following inhalation exposure. If systemic exposures following expected oral and dermal exposure are less than for inhalation, for which there are available exposure-response data, some consideration should be given to limiting toxicity studies. He suggested that studies on immunotoxicity should include doses in the range of human exposures to ensure the relevance of any findings and their extrapolation to actual exposures. Dr. Bucher asked Dr. Teeguarden how he would rank the significance of the proposed studies if there were significant exposures to silica flour. Dr. Teeguarden said if the NTP's surveys suggest significant oral and dermal exposures, then he would rank the significance as moderate, at least for the biokinetic studies. If the biokinetic studies indicate only unreasonably high doses would allow silica flour into the gastrointestinal tract or through the skin for systemic exposures, then a decision should be made not to continue with animal studies. But if, at more reasonable exposures, silica flour crosses the gastrointestinal tract and skin, and the bound particles would be retained and have the potential to cause adverse health effects, then the studies would be warranted. Dr. Bucher said the NTP would reconsider the nomination very carefully.

Dr. Riviere thought transdermal absorption of silica that is 1 µm or larger unlikely. The only potential problem with silica would be skin irritation at hair follicles, which would not lead to systemic absorption. He gave a low priority to studies of dermal toxicity.

F. Research Concept for Valerian

a. Presentation

Dr. Michael DeVito, NIEHS/NTP, presented the research concept for valerian, which most commonly refers to extracts of the underground rhizomes and roots from several subspecies of *Valerian officinalis*. Valerian has been used for neurological and/or psychological ailments, insomnia, mood disorders, anxiety, psychological stress, menstrual cramps and menopausal symptoms. It was nominated by the NIEHS due to limited toxicological data and its prevalence in a number of products as the main constituent or in combination with other herbs (it is the 11th top selling botanical dietary substance). There are at least 150 different constituents in valerian including sesquiterpenoids, valepotriates, and valeric acid. Exposure to valerian is widespread, and potentially higher among young women. Some clinical data indicate that valerian has sedative activity, improves sleep quality, and does not impair or improve psychomotor or cognitive abilities; however, the studies were of short duration and had no follow up. The limited toxicological data in rodents show increased sleep time. Developmental studies showed only minor changes in ossification. There is *in vivo*

evidence of DNA damage in a 90-day male mouse study. No chronic toxicity studies are available and one study reported a LD_{50} of 1.5 g/kg or higher

Key issues in the project are extensive human exposure and limited toxicological characterization. Some testing challenges include determining which preparation to evaluate. There is little consistency in the recommended doses and constituents on valerian labels and there is a very broad range of preparations. The NTP would collaborate with the FDA to determine the valerian species used in modern valerian preparations and establish a consensus preparation. A tiered approach would be used: (Tier 1) chemical characterization to compare extraction procedures based on valerenic acid content, (Tier 2) *in vitro* screens to assess neuronal firing and mutagenicity, and (Tier 3) sufficient similarity studies to find mixtures that are useful to test. A second tiered approach, if warranted, would consist of (Tier 1) repeat dose studies, (Tier 2) reproductive and developmental studies, and (Tier 3) chronic studies.

Significance of proposed research is that it would provide toxicological data to enable understanding of toxicity of valerian root and its constituents and it would further provide data for developing regulatory policy for valerian dietary supplements and herbal preparations.

b. Public Comments

Mr. Manuppello, PETA, said valerian has been used to treat digestive and urinary tract problems for over 1000 years and is also widely used as a mild sedative and sleep aid. The German Commission E recommends valerian for use in the management of restlessness and nervous disturbances of sleep. No serious adverse effects were reported in a 2007 review of 37 clinical trials. The most common side effects were mild dizziness, headache, drowsiness, or gastrointestinal discomfort. The authors concluded that the studies reviewed support the safety of valerian. The NTP expresses concern regarding potential adverse developmental and reproductive effects, citing two studies that suggest valerian extracts retard ossification in rats and mice following developmental exposures. PETA believed the concerns are overstated. Only Tufik *et al.* (1994), using a mixture of valepotriates, reported statistically significant increases in the number of retarded ossifications, at doses 10 to 20 times greater than those prescribed for humans. There are questions about the bioavailability of valepotriates. No additional details of these observations were presented. In a more extensive study using a valerian extract in ethanol, Yao *et al.* (2007) reported no significant difference in the number of ossified sternebrae in fetuses from any of the treatment groups at doses up to 65 times greater than the normal human dose. While the authors observed differences in the number of ossified metacarpals, similar differences were also found in the ethanol control group compared to the water control group. In addition, no significant differences were found in the mean number of implantations per dam, corpora lutea per dam, live fetuses per litter, total number of resorptions, or percentage pre-implantation loss compared to ethanol or water control groups, and no external malformations were observed. The authors concluded that valerian extract did not have any significant adverse effects on fertility or embryo development.

The NTP also states that increases in the frequency of micronuclei in polychromatic erythrocytes observed in mice receiving a valerian extract suggest that some of the constituents may be genotoxic. In fact, these increases were only observed by Al-Majed *et al.* (2006) at the highest dose, about 40 times greater than the recommended human dose. The authors acknowledge that the depletion of testicular nucleic acids observed at higher doses was not dose-dependent and that their data contradict a substantial body of literature showing that many of valerian's constituents are found to be antioxidants known to protect against genotoxicity and/or carcinogenicity.

The NTP states that there are six case reports of hepatotoxicity, but since these patients were taking mixtures of herbal supplements containing valerian and other herbs, the contribution of valerian to the hepatotoxicity is uncertain. In its review document, the NTP notes that researchers now believe that germander, a plant from the mint family and also believed to be present in the herbal medicines, caused the liver damage. In several recent reviews and meta-analyses of the efficacy of valerian involving more than 1000 patients, no occurrence of hepatotoxicity was reported. Also, in a review of 23 cases of self-poisoning with a valerian-containing insomnia treatment commercially available in Hong Kong, Chan *et al*., (1995) found no clinical evidence of acute hepatitis and no evidence of subclinical liver damage in 12 patients on whom routine liver function tests were performed.

Regarding the proposed approach, PETA is concerned that uterotrophic assays could be employed in choosing the test material. The only evidence for valerian's estrogenic activity appears to be from a survey of 17 different plant extracts that produced conflicting results for transcriptional activity and receptor binding *in vitro* (Overck et al., 2008). The proposed *in vitro* screens to evaluate estrogenic activity should be regarded as definitive and negative results should not be taken as a compelling reason to try uterotrophic assays. In addition, an *in vivo* micronucleus assay is proposed in the first tier of toxicity testing. Since conflicting *in vitro* mutagenicity results have been reported, these *in vitro* tests should be repeated and *in vitro* tests should also be employed to investigate valerian's potential to induce chromosomal aberrations in the first tier of testing.

Dr. Teeguarden asked if the meta-analysis was designed to test for hepatotoxicity. He said there is a difference between there being little or no evidence of toxicity/adverse effects and sufficient evidence of safety. He asked Mr. Manuppello which of the cases exists for valerian in the studies described in the comments. Mr. Manuppello said there was no evidence of adverse effects, and PETA is unclear what the reason for concern is considering the long history of safe use of valerian.

c. BSC Discussion

Dr. Carney, first lead discussant, said the rationale was clearly driven by exposure, especially potentially high does in women of childbearing age. The clinical studies done on valerian were not designed to determine toxicity. Assessment of reproduction and development in animals is limited to two studies, both of which are extremely limited in design and do not meet regulatory guideline standards. Teratology cannot be assessed

with the small group sizes used in the studies. The proposal is consistent with the mission and goals of the NTP, and is a great opportunity for the NTP to lead the way on the science of mixtures. He was supportive of the sufficient similarity studies and the neuronal firing screens. He said what is done in the early stages of the project would be very important.

Dr. Faustman, second lead discussant, concurred with Dr. Carney's comments and expressed concern for the potential high exposures of young women to valerian. She said getting the first steps right would be critical and suggested laying out an extraction plan that describes fractionation and assays to be conducted in a “decision like” format. She expressed concern that using only γ-aminobutyric acid screens would cause the NTP to miss other endpoints. There is *in vitro* work on valproic acid structure activity relationships, which could be built upon. Dr. Faustman was unclear about what characteristics would move specific fractions to the toxicity testing in specific aim 2 and suggested more information on the scope of aim 2.

Dr. DeVito agreed that phase 1 is critical. The NTP can potentially use a developmental neurotoxicity screen that the EPA is developing to evaluate neurite outgrowth. The structure activity relationships for mutagenicity may be completely different than those for neuroactive components, so the NTP may study multiple extracts depending upon that difference. With sufficient similarity methods it may be possible to narrow down the number of mixtures that would be tested *in vivo*. Drs. Faustman and DeVito discussed the teratogenic and neurodevelopmental effects of valproic acid relative to screens for valerian.

Dr. Riviere suggested doing the initial screens on two or three mixtures because sufficient similarity would not show similarity between certain extracts. Dr. DeVito explained that more than three mixtures would be used in the studies because there would be additional complexities using acidified and non-acidified extracts. Dr. Bunton asked about receptor binding assays. Dr. DeVito said there is evidence that valerian has actions with γ-aminobutyric acid, adenosine, and serotonin. The NTP could screen all receptors, but a potential outcome is different structure activity relationships or potencies from each assay. Screening neuron firing would be a more useful first screen to ensure that no molecular mechanisms are ignored. Dr. Bunton suggested using a more general receptor binding assay to get a more global perspective.

IX. NTP’s Dietary Supplement and Herbal Medicines Initiative

A. Presentation

Dr. Nigel Walker, Deputy Director for Science, NIEHS/NTP, said Congress defined the term "dietary supplement" in the Dietary Supplement Health and Education Act (DSHEA) of 1994. A dietary supplement is a product taken by mouth that contains a "dietary ingredient" intended to supplement the diet. In 2006, total sales of dietary supplements were $22.1 billion and U.S. herbal dietary supplement sales was $4.59 billion. Use of dietary supplements is widespread in the United States; estimates in 2008 were one in nine children and four in ten adults used dietary supplements. Much of this use is for wellness, rather than for treating specific ailments, which the NTP

considers important in terms of potential chronic usage. The NTP initiated work on dietary supplements in 1998 with a workshop that recommended more research, standardization of product ingredients by industry, increased consumer education, identification of herb-drug and herb-herb interactions, and research on risk to sensitive subpopulations. The NTP identified important, special considerations for herbals when compared to other test articles: (1) which product should be studied in terms of different plant species/plant parts and multiple formulations; (2) possible variation in sources due to different growing, harvesting, and processing conditions; and (3) characterization issues such as differing physical and chemical characteristics, analytical techniques, and study requirements.

The general approach to characterizing herbal-based dietary supplements includes identifying known active agents from the literature, chemical fingerprinting, evaluating with mass spectroscopy, and identifying nutritional constituents. There is differing lot-to-lot variability in many of the dietary supplements. The NTP studies a variety of herbal use categories including multipurpose (e.g., aloe vera), women's health (e.g., black cohosh), cancer chemopreventives (e.g., green tea extract), anti-aging/wellness (e.g., Ginko), and weight loss/sports aids (e.g., Garcinia, which often have rapidly changing formulation)

The NTP's rationale for selection of dietary supplements includes size of the population using a given supplement, use pattern, and potential biological activity. The NTP completed five technical reports for dietary supplements in 2009. In the past year, there has been increased coordination with FDA on moving forward and prioritizing dietary supplement studies. An interagency coordination group was established comprised of NIEHS/NTP, NIH Office of Dietary Supplements, FDA/Center for Food Safety and Applied Nutrition (CFSAN), and FDA/NCTR. The aim of the group is prioritization of agents for study and a more coordinated effort on dietary supplements. CFSAN receives many pre-market notifications on new dietary ingredients. The group is able to identify potentially important data and to design and conduct additional studies to expand mechanistic and translational understanding.

Future directions include (1) chemical and biologically based screening and prioritization using short term in vivo assays and high-density in vitro assays; (2) context on data from NTP studies for representativeness of test article to other formulations, exposure extrapolation, and mode of action studies; and (3) potential human studies with the NIEHS Clinical Research Unit. Dr. Walker closed by saying the NTP would be pursuing increased coordination to ensure obtaining the most needed information to inform public health decision-making.

B. BSC Discussion

Dr. Faustman said it was exciting to pull all the lessons learned together. She encouraged the NTP to disseminate the information about dietary supplements more broadly due to the size of the population affected and its economic impact. Dr. Bucher said some of the information coming out shortly would get the attention of Congress. Dr. Faustman said it would be helpful to put the information in a context where different

options for decision-making could be presented. Dr. Howard said the FDA cannot petition the Congress to change laws. There is great enthusiasm at the FDA for the dietary supplement working group. As the group grows, it will draw in more organizations involved in risk assessment and natural products. The dietary supplement industry is very aware of what the NTP and working group are doing, and it would have an impact on the way the industry handles toxicology in the future. Dr. Faustman said the National Center for Complementary and Alternative Medicine is missing from the working group. Dr. Howard said they would be at future meetings, and Dr. Bucher said they have been involved in the nomination phases. Dr. Nagarkatti suggested that the NTP study other complementary and alternative medicines, e.g., homeopathic medicine. Dr. Faustman said another consideration is the unique exposure to herbal compounds, e.g., in a sauna or skin wrap.

X. Conclusion

Dr. Bucher thanked Dr. Portier for chairing the meeting, the BSC for their comments, and the NTP staff for organizing the meeting. Dr. Shane thanked the BSC for their participation.

XI. Adjournment

Dr. Portier adjourned the meeting at 12:45 PM.

Review of the NTP Host Susceptibility Branch Research and Testing Program

National Toxicology Program

Board of Scientific Counselors

Meeting Report December 9 – 10, 2009

The BSC evaluated the NTP Host Susceptibility Branch (HSB) Research and Testing Program in a public meeting of the BSC on December 9 - 10, 2009 at the National Institute of Environmental Sciences, Research Triangle Park, NC. Four BSC members, Dr. Diane Robins lead, University of Michigan Medical School; Dr. Janan Eppig, Jackson Laboratory; Dr. Stephen Looney, Medical College of Georgia; and Dr. Raymond Novak, Wayne State University, served as primary reviewers of the HSB Program. Dr. David Threadgill, North Carolina State University, served as an *ad hoc* reviewer. Dr. Lori White, NTP Executive Secretary, served as rapporteur for the review.

The BSC was provided with background materials for the review in November 2009. The BSC was asked to review the eight projects in the HSB and then address the charge:

Review and evaluate the research activities of the Host Susceptibility Branch related to the development of an NTP program in host susceptibility.

Background and Overview of HSB Program

Dr. John French, NIEHS, acting chief, HSB, provided an overview of the HSB Program and described the evolution of the Program at the NTP. The HSB mission is to develop and test genetically modified and diverse animal models to determine (a) the range of their biological responses to toxic agents of public health importance, (b) the genetic and epigenetic bases for these biological responses, and (c) the mode or mechanistic basis for agent-specific toxicity in order to improve extrapolation across species. The primary aim of research carried out by the HSB is to develop new laboratory models and testing protocols for hazard identification and risk assessment. The models are intended to quantitatively capture the range of response between individuals that correlate with individual human susceptibility to toxicity and disease. Individual genetic differences harbored within the human population are believed to be the basis for individual susceptibility to environmental stressors that cause toxicity, including idiosyncratic drug toxicities. The goal of the HSB is to model this genetic diversity in the human population using genetically diverse laboratory animals for *in vivo* and *in vitro* toxicology studies.

BSC Discussion

BSC suggested highlighting "dose" in discussions of extrapolation across species in absorption, distribution, metabolism, and excretion (ADME) studies.

Background and Overview of Perlegen Resequencing Project

Dr. Frank Johnson, NIEHS project leader, presented an overview of the NIEHS/NTP-Perlegen Resequencing Project. He explained that in 2004, the NIEHS/NTP embarked on a project to determine the genomic DNA sequence of 15 inbred strains of mice. The project is currently being conducted under contract to Perlegen Sciences in Mountain View, California. The NTP is contemplating how to make best use of project findings to the benefit of toxicology, the testing of chemicals, and public health. Most toxicology testing utilizes very few genotypes, which tacitly assumes genotype makes little or no difference to test outcome. The study utilizes 11 classical mouse strains and 4 wild-derived strains. The results of this project will help identify genes in mice that underlie susceptibility to adverse health conditions such as cancer and heart disease. The results can also help identify genetic factors responsible for variability in the response to toxic agents and help explain why some genotypes may be more susceptible than others to the harmful effects of exposure.

BSC Discussion

The BSC commended the NTP on the project and said its resequencing project models reflect remarkably well what is present in human populations. This has been a great resource that will serve the toxicology community and broader research community as a whole for years to come. These data have been incorporated into a number of databases that allow widespread access. This work has stimulated research in the broader scientific community, so rather than focusing on new sequencing projects, the HSB should focus on applying this information to the other HSB proposed projects.

Project 1: Benzene ADME in Genetically Diverse Mouse Strains

Dr. Michael Cunningham, NIEHS project leader, described HSB's testing of multiple, genetically diverse, mouse inbred strains to determine the variable range of ADME kinetic parameters using benzene, which has been extensively studied in animals and humans. The hypothesis is that genetic variation inherent in multiple inbred mouse strains will show significant differences in ADME phenotypes that determine bioavailability and toxicity. Preliminary data analyses have established significant differences among the 18 strains, confirming the selection of those strains for in-depth investigation for tissue metabolite distribution. At present, the benzene ADME studies conducted in females of these strains are undergoing final evaluation and quality control for in-depth statistical analysis. If required to improve power for haplotype association mapping (HAM) and statistical analysis, additional

strains of inbred mice would be investigated.

BSC Discussion

This is an exciting project that not only meets the prescribed goals of the HSB, but also represents cutting edge research with substantial potential for application to human populations from multiple perspectives. Certainly, accomplishing the stated goal of this project will directly address each of the overall goals of the HSB program.

As these data emerge, it will be important to consider how the data can be applied to help develop models of genetic diversity in human populations

This project has a clear scientific and public health impact. Understanding the role of genetic differences in altering toxicant (i.e., benzene) ADME parameters will provide valuable information across strains, as well as provide seminal information on the genetic basis for altered toxicant ADME and PK/PD models.

The validity of the approach is compelling and the rationale underlying the project is well defined and strongly supported by preliminary data.

The approach was well conceived initially and continues to evolve in the appropriate direction. Additional modeling of the PK/PD data to examine whether genetics alters the number of compartments will provide seminal information regarding the significant impact on animal and human population PK/PD.

Preliminary data have been obtained and are exceptionally strong and highly supportive of the goals of the project. Integration of ADME data with hepatic enzyme expression in the strains / genders examined will further extend the strength of this research.

Future extensions of this research are well conceived and strongly encouraged. Additional studies that may be entertained as part of this research project include the comparison of benzene inhalation exposure ADME with that obtained from the oral administration of benzene and modeling of the PK/PD data.

Careful consideration must be given to the addition of strains and the resulting effect on statistical power. This work is highly dependent on appropriate statistical analyses that will account for the unique genetic structure of these strains. It is not clear that power for the HAM and the statistical analyses will be increased simply by adding strains of mice, since the power of any statistical test is a function of both the within and between variability that the new strains add to the analysis. Adding strains, especially those without more independent genetic compositions, could actually decrease power.

It would be helpful if a timeline could be provided for each of the projects so that the BSC can

adequately assess progress. Specific milestone in addition to a general timeline would be helpful.

The BSC discussed HSB's use of the one-compartment model, and agreed it may be useful to consider adding a multi-compartment model.

Project 2: Benzene ADME Phenotype and Haplotype Association Analyses

Dr. French, project leader, explained that data from Project 1 is being used to perform haplotype association mapping (HAM) with ADME kinetic parameters (AUC, Cmax, Tmax, or CL_F) and to determine the optimal ADME study design for quantitative genetic analysis. The hypothesis is that individual strain genetic variation will result in significant differences in benzene-induced ADME kinetic parameters across mice strains that reflect variation and orthologous genes observed in the human population. Preliminary results show significant strain variation. The kinetic parameter data obtained from the males of the 18 sequenced strains are currently being analyzed by HAM to determine if quantitative trait loci (QTL) can be detected and candidate genetic variants identified for functional validation.

BSC Discussion

The project to investigate host variation in ADME phenotypes is critical to understanding how genetic variation and multi-strain panels may be exploited to better understand variation in response to toxicants. This project seems highly appropriate for the HSB as part of its mission to investigate the causes of toxicant response variation.

The project appears to be equally central to extrapolating exposure research to human populations. Understanding whether variation in ADME phenotypes is due to phase 1 or 2 drug metabolizing enzyme genes or other as yet unidentified genes is important.

The hypothesis for Project 2 states, "Individual strain genetic variation will result in significant differences in benzene induced ADME kinetic parameters in multiple strains of mice that reflect variation and orthologous genes observed in the human population." It is essential to quantify the association between the differences seen in benzene ADME in already examined strains of mice and the corresponding variation observed in diverse human populations.

The approach being pursued is appropriate. However, like all current approaches using HAM analysis with existing inbred mouse panels, there are many caveats. Recent data have indicated that the population structure of existing inbred mice is suboptimal for efficient HAM analysis. Similarly, if the causative genetic variants are due to allele combinations, most potential pair-wise allele combinations do not exist in extant inbred strains.

It is certainly appropriate to extend the scope of this project beyond the initial test agent benzene

to a cardiotoxic agent. The idea of "serial sampling of small volumes of blood over time at different exposure levels" as an alternative to discrete sampling of small numbers of animals at individual time points, is certainly well conceived and could potentially provide more "bang for the buck" by eliminating the effect that between-animal variability can have on the statistical comparisons of the time points. However, care must be taken to properly account for the within-animal variation that naturally results when serial sampling is performed (the repeated measures effects).

The BSC said it was important to recognize that diet may interact with exposures to influence phenotypes, directly or indirectly, in a strain-dependent manner. Hence, the relevance to human metabolic state needs to be considered.

Project 3: Benzene Low Dose Inhalation Induced Hematotoxicity and Genotoxicity Phenotypes and Haplotype Association Analyses

Dr. French, project leader, explained that data from Project 1 is also being used to design a low dose benzene inhalation study for quantitative measurement of hematotoxicity and genotoxicity, and to determine the association between benzene ADME kinetics and inhalation toxicity. The hypothesis is that genetic variants (SNP and/or structural) between strains determine the exposure-level-dependent, tissue-specific metabolism of benzene and tissue-specific benzene toxicity. To estimate benzene hematopoietic toxicity and circulating blood cells in each exposure group, HSB will phenotype inbred strains of male mice exposed to low levels of inhaled benzene for 28 days and examine the effect of genetic variation on hematopoietic toxicity and genotoxicity. The C57BL/6, C3H/HeJ, and their F1 hybrids are currently being exposed to validate benzene vapor generation, target exposure levels, and practice collecting tissues for analysis prior to initiation of investigation of 34 inbred mouse strain as a model for the human population.

BSC Discussion

This project clearly addresses the goals of the HSB through the use of genetically diverse and genetically modified animal models, examination of the genetic and epigenetic basis for the variable biological response to toxic agents, and research on the mechanistic bases for toxicant actions that are highly conserved.

This project has an important and relevant health impact in establishing heritable determinants that affect bioavailability and toxicity. Thus, the project addresses fundamental issues of individual susceptibility and, when coupled with other project studies, can address this important issue.

The validity of the proposed research is established based on human exposure to benzene and resulting public health concerns. The rationale is well developed, as inhalation exposure is the

primary route of human exposure.

The approach was well conceived and justified. The analysis of the data will be the most challenging aspect of the proposed research project.

The aims of the project are logical and the information will be of significant value in understanding individual susceptibility to the hematopoietic toxicity and leukemia risk associated with benzene.

Future extensions of this project to other strains, for which there are genetic predispositions to leukemia, should be considered.

A potential future extension of this project, beyond additional strains, might include an examination of epigenetic effects associated with benzene exposure.

The BSC suggested comparing ADME characteristics from inhalation and oral exposures. The BSC felt that this could be one strength of the proposal in that it parallels a number of molecular epidemiology studies of benzene-exposed workers.

Project 4: Studies to Identify Environmental Cardiotoxins and Susceptibilities to Cardiotoxicity

Dr. June Dunnick, NIEHS, project co-leader, described the HSB studies to develop a multiple strain mouse model for the identification of environmental cardiotoxic agents and susceptibility to cardiotoxicity. Components of this program include identification of sensitive and specific biomarkers of cardiotoxicity, identification of more sensitive models for cardiotoxicity testing, and discovery of genetic determinants of chemical-induced cardiotoxicity (CICT). The HSB will expose 34 strains of mice to two mechanistically distinct cardiotoxins, bis(2-chloroethoxy) methane (CEM) and ephedrine/caffeine (E/C), and quantify the extent of cardiotoxicity using serum biomarkers. HSB has found cardiac phenotypic variation in three strains of male mice (C3H/HeJ, C57BL/6J, and B6C3F1/J) that will help identify candidate genes for heritable cardiomyopathy. Candidate genes (Prkaa2, Calr3, and Il15) for cardiomyopathy were identified in C3H/HeJ mice.

BSC Discussion

The BSC suggested using a greater variety of methods to filter and prioritize candidate biomarkers and detect modifier loci.

The initial project is encouraging and provides a solid proof of concept. The larger project, including additional strains, as outlined, should yield some very interesting cardiomyopathy susceptibility genes and produce useful biomarkers.

Project 5: Aging and Spontaneous Disease Phenotypes in Selected Inbred Strains

Dr. French, project leader, described the HSB efforts to establish a benchmark reference database on aging and disease in 10 genetically diverse mouse inbred strains (129/SvlmJ, A/J, C3H/HeJ, C57BL/6J, CAST/EiJ, NOD.B10H2^{b}/LtJ, NZO/HiLtJ, PWK/PhJ, WSB/EIJ, and B6C3F1/J). HSB investigators hypothesize that the genetic diversity in the 10 strains selected will show a significant age-related range of spontaneous disease and functional phenotypes that will aid selection of strains for toxicology and carcinogenesis research and testing. Aging and survival analysis (spontaneous disease) studies are in progress. Three protocols for identifying and developing biomarkers at specific life stages, and functional analysis of cardiopulmonary functions are in progress and under peer-review.

BSC Discussion

Developing the public benchmark reference database described in this project will certainly aid scientists in selecting the most suitable strains for chemical toxicology and disease studies for extrapolation across species.

The rationale for the choice of the 10 inbred strains was clearly spelled out. More detail on the choice of statistical methods and power calculations would be helpful. The discussion of the advantages associated with using different paradigms for comparison of data was somewhat confusing. The statistical power resulting from a two-factor (exposure x strain) design depends on the between and within variance components. If significant interaction between exposure and strain is present, the power of the test to compare the exposure groups will tend to be decreased in the two-factor design since simple effects for exposure (based only on the 10 animals in each strain) would be tested rather than main effects (which would be based on 80 animals per exposure group, ignoring strain). Pilot data are needed for determining if it is likely that there is statistically significant interaction between exposure and strain.

Project 6: Development of a Short Term Cancer Bioassay using Multiple

TRP53 Haploinsufficient F1 Inbred Strains

Dr. French, project leader, explained that the database described in Project 5 will be used to prepare for multiple-strain toxicology and carcinogenesis studies and to develop and conduct a short-term cancer bioassay using multiple p53 haploinsufficient F1 inbred strains. The hypothesis is that (1) tumor spectrum, prevalence, and latency, (2) transcript or metabolomic expression profiles, or (3) expression profiles (corroborated by copy number variation) will segregate according to the haplotype of p53 haploinsufficient F1 hybrid isogenic lines selected on the basis of genetic variation in DSB repair genes. Preliminary data in C3B6F1 p53 haploinsufficient mice have shown a non-random allele-specific loss associated with Melm3, Trp53, and Rad51c genes. To date, survival and tumor phenotypes observed by histopathology

are significantly different among the four completed F1 hybrid mice studies. Information from these studies will allow NTP to select inbred strains to perform pre-chronic and chronic multi-strain toxicity and disease studies with NTP nominated chemicals.

BSC Discussion

The main aim of this project is to develop and test a rapid and predictive model for non-random loss of heterozygosity associated with tumor suppressor genes with sufficient genetic diversity to model the human population. The goal is to enable identification of a mode of action for hazard characterization, reduce potential for false positives and false negative, and aid extrapolation of risk to human populations.

Preliminary data in C3B6F1 p53 haploinsufficient mice have shown a non-random allele specific loss associated with Melm3, Trp53, and Rad51c genes. Four of the six strains have been completed and the remaining two will be completed by December 2009. To date, survival and tumor phenotypes observed by histopathology are significantly different between the four completed F1 hybrid mice studies. Specific information on benchmark accomplishments, future goals, and timelines would be useful.

Project 7: Alkylanilines Class Study

Dr. Scott Auerbach, NIEHS, project leader, described the Alkylanilines Class Study, which has the goals of quantifying the degree to which genetic variation influences the genotoxicity of select alkylanilines and identifying alkylaniline genotoxicity quantitative trait loci. A subset of alkylanilines was nominated to NTP for toxicological characterization because of the potential for widespread human exposure, the limited availability of published toxicological data for this subclass of alkyl-substituted anilines, and their structural similarities to two known animal carcinogens. Initial studies to evaluate all 14 alkylanilines for genotoxic potency using an AS52 cell assay are currently being designed. The studies have the potential to identify currently unknown genetic loci that modify the carcinogenic risk associated with alkylaniline exposure.

BSC Discussion

The merit of the proposed research is high and should provide valuable information on the basis for alkylaniline-induced toxicity.

The project has a clear scientific goal and public health impact given the level of human exposure to alkylaniline, their genotoxicity and their role in human cancer.

The validity of the proposed research is based on human exposure and cancer and the rationale for the research project is well established.

The approach is well conceived, although the proposed studies with primary cultured

hepatocytes should consider the variability of the primary cultured hepatocytes and incorporate some measure by which the various cell culture preparations may be deemed viable (e.g., CYP1A2 levels; response to H_20_2 for the Comet assay as well as repair time) for use in examining DNA adduct levels. Widely disparate levels of the enzymes required for alkylaniline metabolic activation and DNA adduct formation may be misleading when associated with the genetics of the strains being examined. Some primary cell preparations are not metabolically viable, even though there is no evidence of aberrant morphology or cell death. The media, levels of insulin and other factors (e.g., time prior to treatment) will be extremely important in this research, as the level of enzyme activity will be affected by cell culture conditions and will reflect the overall functional capacity of the cell and the DNA damage. Since genetics may also govern differences in DNA repair, the Comet assay could be used to assay the time required for cellular DNA repair.

The aims of the project are straightforward and should be readily accomplished with the appropriate controls and conditions.

Future extensions to examination of genetic associations with DNA adducts are appropriate, although hepatic tissue as well as the bladder tissue should be the target of such research.

Some directed studies employing the genetically diverse strains for select alkylaniline ADME studies to examine target organ chemical levels and examination of bladder tissue for DNA adduct should be considered as the difference in organ response may well reside in the initiating enzymes and/or repair capacities of the cells.

Project 8: Impact of Sex and Strain on the Performance of Genomic Signatures of Hepatocarcinogenesis

Dr. Auerbach, project leader, described the development of studies to determine the impact of sex and strain on the performance of genomic signatures for predicting hepatocarcinogenesis. Male F344/N rats, exposed to 30 distinct chemical treatments, will be used to create a 4-class hepatocarcinogenicity prediction model; the four 4 classes are genotoxic hepatocarcinogen, non-genotoxic hepatocarcinogen, hepatotoxic non-carcinogen, and non-toxic non-carcinogen. The classification accuracy of the model will be tested on female F344/N rats and on male and female Sprague Dawley and Wistar Han rats. Chemical and dose selection are underway and initial studies should be started this fiscal year. HSB will use a similar approach to determine if predictive signatures can accurately classify gene expression changes in mouse.

BSC Discussion

BSC said this research is exciting in that it brings a male/female comparison of carcinogen-mediated differential gene expression into consideration when discussing genetic susceptibility. BSC questioned whether genomic signatures resulting from human hepatocellular carcinoma are

in existence that could be used to compare with the rat. BSC suggested making more direct use of HSB data, which apparently is in press but was not provided for review, making it difficult to ascertain whether a more focused approach might be more appropriate. It is not clear how the pronounced sexual dimorphism in rodent liver gene expression is taken into account in this study.

BSC was concerned that the outbred rat strains lacked the genetic homogeneity of mice, a positive aspect of the previous projects. Thus, it is difficult to link gene expression signatures with terminal cancer phenotypes.

Summary

In summary, the HSB review panel was highly enthusiastic about the program. The review panel expressed the view that the program as a whole has made enormous progress in accomplishing its mandated goals. Applying contemporary genetic approaches to address current limitations in toxicological studies is highly relevant and directly applicable to human populations, in which such experiments cannot be conducted. This unique program is poised to take maximal advantage of its activities and results to advance human health.