National Toxicology Program
Board of Scientific Counselors

November 20-21, 2008

National Institute of Environmental Health Sciences
Research Triangle Park, NC

Summary Minutes

Table of Contents

I. Attendees

Members in Attendance:
Tracie Bunton, Eicarte LLC
Edward Carney, The Dow Chemical Company
Russell Cattley, Amgen
Kenny Crump, Louisiana Technical University
George Friedman-Jiménez , New York University School of Medicine
William Janzen, University of North Carolina
Nancy Kerkvliet, Oregon State University, (chair)
Jon Mirsalis, SRI International
Raymond Novak, Wayne State University
Michael Pino, Sanofi-Aventis (on the telephone for part of the NTP Nominations and Concepts and the NTP BSC Working Group Report on Criteria for Evaluating Outcomes in Immunotoxicology Studies and Reproductive and Developmental Toxicology Studies)
Kenneth Portier, American Cancer Society
Jim Riviere, North Carolina State University
Diane Robins, University of Michigan Medical School
Keith Soper, Merck & Company
David Wegman, University of Massachusetts, Lowell

Members not in attendance:
Katharine Hammond, University of California Berkeley
Gail McCarver, Medical College of Wisconsin

***Ad Hoc* Member**
John Vandenbergh, North Carolina State University

National Institute of Environmental Health Sciences (NIEHS) Staff

Scott Auerbach	Barbara Shane
Eddie Ball	Michael Shelby
Chad Blystone	Keith Shockley
John Bucher	Cynthia Smith
Rajendra Chhabra	Jennifer Smith
Gregg Dinse	William Stokes
Paul Foster	Matthew Stout
Dori Germolec	Kristina Thayer
Richard Irwin	Raymond Tice
JoAnn Lewis	Molly Vallant
Robin Mackar	Nigel Walker
David Malarkey	Lori White
Scott Masten	Samuel Wilson
Christopher Portier	Kristine Witt
Joseph Roycroft	Mary Wolfe

Other Federal Agency Staff
Paul Howard, Food and Drug Administration (FDA)/ National Center for Toxicological Research (NCTR)
Mark Toraason, Centers For Disease Control and Prevention (CDC)/National Institute for Occupational Safety and Health (NIOSH)

Public
Neepa Choksi, Integrated Laboratory Systems, Inc.
Charles Hebert, Southern Research Institute
Claudine Gregorio, Integrated Laboratory Systems, Inc.
Marc Jackson, Integrated Laboratory Systems, Inc.
Joseph Manuppello, People for the Ethical Treatment of Animals
Glen Marre, Jr., EPL, Inc.
Rodney Milton, EPL, Inc.
Jacqueline Rams, Southern Research Institute
Ivan Rusyn, University of North Carolina – Chapel Hill
Carol Swart, Integrated Laboratory Systems, Inc.
Alan Staple, RTI International
Michael Waters, Integrated Laboratory Systems, Inc.

II. Introductions and Welcome

The National Toxicology Program (NTP) Board of Scientific Counselors (BSC) met on November 20-21, 2008, at the Rodbell Auditorium, National Institute of Environmental Sciences, Research Triangle Park, North Carolina. Dr. Nancy Kerkvliet served as chair for Dr. Gail McCarver in her absence. She 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.

Dr. Samuel Wilson, Acting Director of the NIEHS and NTP, welcomed the BSC members and expressed his gratitude to them for their attendance at the meeting and for their advice to the NTP on its activities. He presented certificates of services to Kenny Crump, Nancy Kerkvliet, Jon Mirsalis, and Keith Soper whose term of service on the BSC would end in December 2008 and thanked them for their dedication and advice. The terms for Drs. Gail McCarver and Katharine Hammond also end in December.

III. NTP Update

Dr. Bucher, NTP Associate Director, NIEHS, welcomed the members of the BSC and outlined the latest activities of the NTP.

NIH Research Festival
NTP scientists made presentations at the recent NIH research festival in October. This annual event accepts proposals for presentations from all the NIH institutes and centers and approximately 20 topics are accepted for presentation. This year was the second time

in three years that the NTP was invited to participate. This symposium highlighted some of the activities of the Host Susceptibility Branch. The overarching theme for the talks was "Genetic Susceptibility – The Link between Environmental Exposure and Human Disease." The NIEHS and NTP have been focusing efforts on an exposure biology program and trying to understand exposures related to environmentally induced diseases while other institutes have been focusing on human genetics to identify the genetic variance that might provide susceptibility to disease. At the last BSC meeting, Dr. John (Jef) French discussed the host susceptibility program in which murine genetic information and SNP patterns might provide an understanding of the range of susceptibilities in environmentally induced diseases in humans. The symposium featured several investigators from the NCI who presented talks on breast cancer and nervous system tumors, Dr. June Dunnick who discussed environmental and genetic factors in cardiac disease models, and Dr. French who presented his research on DNA strand breaks and their relationship to the development of cancer.

Workshop to Discuss the Pathology of the Reproductive Tract
The NTP held a 2-day workshop on the pathology of the reproductive tract. This workshop is the second in a series to address the pathology of non-cancer endpoints and develop consistent terminology for describing the findings. The first workshop dealt with the pathology of the immune system.

Vendor Meeting on Assays for High Throughput Screening
The NTP held a meeting for the high throughput screening initiative and invited companies to provide information on the kinds of commercial assays they have developed for biological targets for drug development. The NTP also asked the companies to provide information about their products, specifically in the context of how they could be used for toxicity screening. Twenty-seven companies presented at the meeting and there was excellent exchange as they also realized the value of their products not only for drug discovery but also for toxicology. The meeting was informative for the NTP to understand the kinds of assay systems available commercially that might be applicable to high throughput screening systems. The companies became aware of how they might develop or adapt their assays to be useful in the evaluation of toxicological pathways.

Center for the Evaluation of Risks to Human Reproduction (CERHR)
Dr. Bucher provided an update on CERHR activities since the BSC meeting in June when the draft NTP Brief on Bisphenol A was peer reviewed. The final NTP-CERHR Monograph on Bisphenol A was issued September 3, 2008. It contains the NTP Brief, the CERHR Expert Panel Report, and the Website links for the public comments on the draft brief and for the BSC peer review report.

The NTP believes that addressing data gaps for BPA is important; however, before proceeding with studies to address research needs, the NTP wanted to understand what studies were underway or being planned on BPA by the extramural community. The NTP and NIEHS Division of Extramural Research and Training issued a Request for Information on October 21 with a closing date of December 1, 2008, to ascertain the

types of studies under investigation with BPA by the academic community and industry, specific data needs for priority areas, and suggestions for beneficial research collaborations.

The NTP and the FDA/NCTR have initiated studies to obtain data for constructing PBPK models in rodents and nonhuman primates. Both groups are planning studies to explore the long-term consequences of perinatal exposure to BPA to understand the potential impact to humans of the developmental changes reported in numerous laboratory animal studies. It is hoped that the PBPK models will be able to link information from rodent studies with primate studies, and potentially with human outcomes.

CERHR recently decided to reevaluate the information on genistein and soy formula, primarily due to the length of time that had elapsed between the expert panel meeting on gensitein and soy formula and the large amount of new literature that has been published.

Review of the Criteria for Evaluating Outcomes from Non-cancer Studies
One of the primary activities for this meeting is the review of the draft NTP criteria to evaluate the outcomes of non-cancer studies. Two NTP BSC working groups met, one in August to discuss and recommend changes to the criteria drafted by the NTP for evaluating the outcomes of NTP immunotoxicology studies and a second in September to discuss and recommend changes to the criteria drafted by the NTP for evaluating the outcomes of NTP reproductive and developmental toxicology studies. The reports prepared by the working groups will be reviewed by the BSC at this meeting. The final NTP criteria will be explained to the larger toxicology community at a special session scheduled for the Society of Toxicology (SOT) meeting in Baltimore in March 2009. Initially, the criteria will not be static, but after a period of use, the NTP will assess their appropriateness and make modifications if necessary.

Future Meetings
Progress is being made on the Report on Carcinogens (RoC). Expert panel meetings have been held for captafol, o-nitrotoluene, aristocholic acids, riddelliine, and styrene. An upcoming meeting on cobalt-tungsten carbide powders and hard metals will be held in December. The first five draft substance profiles containing the NTP's draft policy decision to list or not list the candidate substance in the 12th RoC will be brought to the BSC on February 24, 2009, for peer review. The Technical Reports Review Subcommittee meeting will follow on February 25, 2009, where six draft NTP Technical Reports will be peer reviewed.

Dr. Bucher pointed out that the NTP would be 30 years old on November 15. In celebration there would be a cake at the afternoon break.

IV. Nominations to the Testing Program

Dr. Scott Masten, NIEHS, provided background information on the source of the nominations to the NTP testing program and how they are developed into research projects. There are multiple levels of review of nominations to determine merit and

priority. Substances are selected for study based on their known or anticipated human exposure, production level, suspicion of toxicity based on structure or existing health effects data, availability of adequate toxicological data, public concern, and the utility of additional studies for public health decision-making. He briefly outlined the review process. The first level is the NTP Interagency Committee for Chemical Evaluation and Coordination (ICCEC) that reviews the nominations and suggests the most appropriate types of studies. Research concepts are prepared by NTP study scientists and reviewed internally before presentation to the BSC. Public comments on the nominations and research concepts are solicited via the Federal Register. The BSC is asked to advise the NTP on the merit and priority of the studies outlined in the research concepts. The NTP revises the research concepts based on the BSC's input. The NTP Executive Committee reviews the revised research concepts, public comments and recommendations from the BSC, and makes a final recommendation on whether to move forward with the proposed study programs.

There are five draft research concepts for review: dimethylamine borane, ethylene glycol-2-ethylhexyl ether, bisphenol AF, β-N-Methylamino-L-alanine**,** triclosan, and hydroxyurea**,** the latter of which studies are not recommended. 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 determine whether sufficient justification is provided for the use of the NTP's testing program resources to carry out the proposed research projects as outlined in the draft research concepts. 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.

1. Dimethylamine borane (DMAB)

a. Presentation

Dr. Dori Germolec, NIEHS, presented the draft research concept for dimethylamine borane (DMAB) that was nominated by NIOSH's Dermatology Cross-sector Program for evaluation of dermal absorption, toxicity, skin sensitization, and possible systemic effects. DMAB is widely used in the manufacture of high-temperature printed electronic circuit boards, thin metal films, semiconductors, and power transistors and as a reducing agent.

Dr. Germolec described an occupational accident affecting four workers in Japan. Three workers were decontaminated soon after the exposure but experienced nausea, vomiting, and gastrointestinal tract effects that resolved in 24 hours. One worker did not immediately decontaminate and experienced severe and persistent neurological symptoms. The NTP postulated that DMAB is absorbed through the skin and targets the nervous system and possibly other organ systems.

There are some LD50 data, but no standard, repeat-dose subchronic or other toxicity

studies for DMAB. The toxicokinetics are unknown following dermal exposure, but in an aqueous environment, DMAB slowly dissociates into boric acid, hydrogen boride, and dimethylamine. Dimethylnitrosamine, a known carcinogen, is formed from DMAB in the GI tract in the presence of a nitrosylating agent.

The first specific aim of the proposed studies is to evaluate the absorption, distribution, metabolism, and excretion (ADME) of DMAB in an *in vitro* human skin model and *in vivo* in rats and mice following dermal, intravenous, and oral exposure. Particular attention will be paid to whether the compound forms dimethylamine or dimethylnitrosamine *in vivo*.

The second aim is to test the genotoxicity of DMAB, and the third specific aim is to evaluate the potential for DMAB to induce dermal irritancy and hypersensitivity. These aims will be conducted in parallel with the ADME studies. Based on the data obtained from the ADME studies, the NTP will decide whether to conduct repeat-dose toxicology studies. If warranted, a 14- and a 90-day study will be conducted using the dermal route of exposure, and a functional observation battery for neurotoxicity will be included in the 90-day study.

The current exposure limits for DMAB are based on dimethylamine and do not take into account whether the borane moiety is toxic. These studies will assess this potential toxicity and will determine whether regulatory limits for DMAB are needed.

Dr. Germolec provided information on some of the initial studies that have been performed on DMAB. She said it was positive in the murine Local Lymph Node Assay and is a dermal irritant.

b. Public Comments

Mr. Joseph Manuppello presented comments prepared by Nancy Douglas on behalf of People for the Ethical Treatment of Animals (PETA). NIOSH based its nomination on DMAB on the possibility that it can cause an allergic response and its systemic toxicity. DMAB is already known to cause eye, skin, and respiratory irritation in humans. If additional dermal absorption data are needed, then validated *in vitro* human skin absorption methods such as the Organization for Economic Co-operation and Development (OECD) test guideline 428 can be used. Workplace and engineering controls as well as protective clothing already limit occupational exposure, and results from the proposed animal tests would not change the safety precautions already in place. A case study reported in 2005 provides evidence that skin sensitization is not a likely outcome of human exposure and no increase in skin sensitization has been reported following years of industrial use.

PETA is concerned about basing a large animal study on a single incident of exposure and using very high concentrations of DMAB for the animal study. There is no indication that the use of DMAB following accepted safety protocols causes long-term neurotoxicity and as indicated in the case report, proper decontamination procedures eliminate this risk. PETA urges the NTP to assign this nomination a low priority. At a

minimum, widely accepted *in vitro* testes on human skin could be used to confirm dermal absorption, and epidemiology studies based on occupational exposure to DMAB over the past 15 years should be incorporated into the test plan.

c. BSC Discussion

Dr. Jim Riviere, a primary discussant, said there is potential for dermal exposure in occupational settings and anecdotal case studies suggest systemic neurotoxicity could occur after dermal exposure. The proposed study is within the NTP's goals. The relationship of dermal DMAB absorption relative to subsequent biotransformation to dimethylnitrosamine is of high scientific interest. Since the hypersensitivity studies have already been conducted, the key question is to determine if dermal absorption and systemic toxicity occur after topical application. This could be addressed first by using an *in vitro* human skin model to ensure that *in vivo* ADME studies are warranted. The remaining studies seem reasonable, but the key issue is the relationship of topical exposure to systemic neurotoxicity, which will require a high transdermal absorptive flux. A rodent model will not provide this information relative to humans, and a human *in vitro* skin model will not provide information that is relevant to humans because it only assess whether absorption occurs.

Dr. Diane Robins, the second discussant, agreed with Dr. Riviere's comments, but she thought the range of testing broader than necessary. The compound is known to be toxic, industry limits exposure of its workers, and only a few people are potentially exposed. Skin testing *in vitro* could be informative and understanding the metabolic breakdown of DMAB would be important, but it is not clear that the results of the other studies would alter the way the compound is currently handled. She gave it a moderate to low priority.

Dr. Jon Mirsalis also agreed with the comments and suggested that the *in vitro* human skin absorption test should be the first aim along with the genetic toxicity studies. He was surprised that a functional observation battery would only be evaluated in the 90-day study. In the one incident with human neurotoxicity, exposure was at a high dose. Dr. Mark Toraason, said NIOSH needs the data from these studies to make skin notations.

Dr. Kenny Crump asked Dr. Toraason what the minimal amount of data NIOSH requires to make skin notations and Dr. Germolec responded on his behalf. She said she is on the NIOSH skin notation work group and was the reviewer for this compound. The work group determines if a compound is a carcinogen, irritant, or skin sensitizer. There was no data on DMAB except for old acute LD50 data.

Dr. George Friedman-Jiménez said DMAB is a well-regulated substance and exposures in the workplace are not a problem. However, he routinely sees workers exposed to well-regulated substances where regulations are not adequately enforced or effective. He does not think that this should be a reason to lower the priority of studying DMAB and human exposure and its effects should be better understood.

2. Ethylene Glycol 2-Ethylhexyl Ether

a. Presentation

Dr. Chad Blystone, NIEHS, outlined the research concept for ethylene glycol 2-ethylhexyl ether (EGEHE), which was nominated by the NIEHS. EGEHE is a solvent used in a variety of products including coatings, paints, lacquers, polishers, and cleaners. Production has been increasing since 1990 and 1-10 million pounds was produced in the United States in 2002. Occupational or non-occupational exposure may be dermal or via inhalation, but no exposure data are available.

EGEHE is structurally related to other well-known glycol ethers and may be converted to the corresponding alkoxyacetic acid, which is responsible for glycol ether's toxicity. The metabolites of short chain glycol ethers, including ethylene glycol monomethyl ether (EGME) and ethylene glycol monoethyl ether (EGEE), cause degeneration of the seminiferous epithelium and are developmental toxicants causing skeletal malformations and lesions in the thymus, bone marrow, and spleen in subchronic studies. There was some evidence of carcinogenicity of ethylene glycol monoethyl ether (EGBE) in male and female mice following inhalation, but no evidence in male rats and equivocal evidence in female rats. IARC concluded that EGBE was not a human carcinogen and classified it as a Group 3 carcinogen.

EGEHE has low acute toxicity. One unpublished subchronic study of male rats exposed for six weeks to EGEHE reported high mortality and degeneration of spermatozoa in the epididymis in the top dose group and some hematoxic effects in the lower dose groups.

The NTP believes EGEHE is worthy of study because (1) it is a high production chemical, structurally related to other known glycol ethers that affect reproduction, development, and hematology; (2) it may yield known or unknown toxic metabolites; and (3) its toxicity after *in utero* or chronic exposure is unknown.

The first specific aim for this research is to characterize ADME, identify known or potential EGEHE metabolites following different routes of exposure, and assess differences between males and females.

The second aim is to evaluate EGEHE's toxicity, especially testicular and hematoxic effects, following subchronic exposure. Developmental toxicity will then be studied using an *in utero* exposure regimen. Genetic toxicity will be evaluated in *in vitro* studies in parallel with the ADME studies. The ADME and subchronic studies can be used to provide guidance on the need for larger and/or more specific studies to evaluate immune toxicity, reproductive toxicity, or chronic toxicity.

b. Public Comments

Mr. Mannupello, PETA, suggested that NTP coordinate its proposed studies with the OECD that has begun to evaluate the toxicity of EGEHE. NIEHS nominated EGEHE based on its widespread use, unknown toxicity profile, and structural similarity to other glycol ethers, whose toxicities have been thoroughly characterized. Scientific evidence indicates that it is unlikely that EGEHE will display a similar toxicity profile to the low molecular weight glycol ethers, and the relevance of observed toxicity to human exposure

is questionable. Although ethylene glycols cause hemolysis in rodent erythrocytes, human erythrocytes are more resistant to hemolysis due to their greater antioxidant capacity. An accepted PBPK model demonstrates that even at saturated vapor concentrations of EGBE, blood concentrations of butoxyacetic acid reached in humans via inhalation will not cause hemolysis and the vapor pressure of EGEHE is lower than that of EGBE. Based on studies from Eastman Kodak, it appears that the skeletal and soft tissue malformation in animals exposed to EGME and EGEE are secondary to maternal toxicity. Simulated inhalation exposure to methoxyacetic acid, the metabolite of EGME, at 5 ppm predicts human blood levels of approximately 16 µM after 8 hours, which is well below the concentration causing adverse effects in pregnant mice or rats.

The increased incidence of liver hemangiosarcomas in male mice and forestomach tumors in female mice following chronic exposure to EGBE is irrelevant to humans, as the former is likely the result of oxidative stress subsequent to red blood cell hemolysis and iron deposition, and the latter to the absence of a forestomach in humans.

The NTP notes the reproductive and developmental effects of glycol ethers are inversely related to alkyl chain length suggesting that EGEHE may not have the same toxic effects reported for the shorter chain glycol ethers. NTP's concern is whether EGEHE is metabolized to shorter chain glycol ethers and shorter chain alkoxyacetic acid metabolites. The NTP proposes metabolism studies of EGEHE in rodents to determine if shorter chain glycol ethers or acetic acids are generated *in vivo*. If this were the case, it would suggest that EGEHE is a testicular toxicant and further studies might not be needed. PETA recommends that NTP first study the metabolism of EGEHE in an *in vitro* hepatocyte cell culture or in mixed cultures of Sertoli and germ cells prior to initiating an *in vivo* study.

c. BSC Discussion

Dr. Tracie Bunton, the primary discussant, said the rationale was clearly articulated that EGEHE is widely used and its chronic effects are unknown. Because, closely related compounds cause reproductive and hematological effects, she is suspicious of EGEHE. However, she questioned NTP's suggestion that EGEHE is a reproductive toxicant based on a subchronic study where two rats out of 10 showed spermatozoa and spermatid abnormalities at a lethal dose.

She thought that the public comments from Eastman Kodak referred to the same study that the NTP cited, but the information provided was different. According to Eastman Kodak, the hematological effects were not of toxicological significance. She concurred with some coordination between the NTP, OECD, and REACH to prevent duplication. She requested that evidence be obtained regarding whether or not true data gaps exist prior to initiating the studies. The proposed metabolism studies in the first tier of testing are appropriate; however, the information presented does not support a comprehensive reproductive toxicity program.

Dr. John Vandenbergh, an *ad hoc* discussant, agreed with Dr. Bunton's comments and said the potential for EGEHE to undergo a similar metabolic pathway to other glycol

ethers needs to be explored to clarify and confirm whether the chain length of the molecule alters the formation of toxic metabolites. OECD's completed studies indicate that EGEHE is unlikely to cause severe toxicological effects. He also agreed that the NTP should coordinate its studies with those at OECD. He assigned EGEHE a low to moderate priority for testing and noted that animal-intensive studies should be initiated only if the metabolic studies indicate possible adverse effects.

Dr. Edward Carney agreed with the comments and added that much is known about glycol ethers especially with regard to structure-activity relationships; this study would determine whether branched chain glycol ethers have similar toxic effects. He believed that EGEHE would be metabolized to an alkoxyacid as it has a primary alcohol moiety and suggested that the studies be done in human and rat hepatocytes. If the production figures are correct, they exceed the trigger for extensive reproductive and developmental toxicity testing under REACH. He suggested that *in vitro* assays could be used to support a "read-across" (category) approach and that this could reduce the need for some or all reproductive and developmental toxicity testing EGEHE in animals. He estimated that EGEHE would fall between the most toxic and least toxic glycol ethers, which would support a category approach.

Dr. Blystone responded that NTP would coordinate with OECD and if OECD identified any needed studies, the NTP would incorporate them into its study design. NTP can incorporate human hepatocytes into the study design. The ADME studies will obtain information on absorption and excretion via inhalation and dermal exposures, and identify possible differences between sexes.

3. Bisphenol AF

a. Presentation

Dr. Matthew Stout, NIEHS, provided background information and the rationale for the nomination of bisphenol AF (BPAF) by the NTP. The NTP Center for the Evaluation of Risk to Human Reproduction (CERHR) recently presented a brief on BPA, in which the NTP expressed "some concern" for effects on the brain and behavior in fetuses, infants, and children and on the prostate gland in adults at current human exposures to BPA. BPAF was nominated for testing following a review of BPA-related compounds. Structurally, BPAF resembles BPA except for the replacement of the six hydrogen atoms in the two methyl side chains with fluorine atoms, suggesting that it could be persistent in the environment. Although production and use patterns suggest that human exposure to BPAF is much lower than to BPA, BPAF may be a more potent synthetic estrogen than BPA.

A key issue for the research program on BPAF is to determine if exposure results in toxicity similar to that of other synthetic estrogens. A related goal is to ascertain whether a structure-activity relationship exists for the class of BPA-related compounds. Three specific aims are proposed to provide a better understanding of the toxicology of BPAF and to test the hypothesis that BPAF produces toxicity characteristic of a synthetic estrogen. The first aim is to conduct a transgenerational study by exposing rats from gestation through sexual maturity to BPAF. The study will assess both standard toxicity endpoints as well as sensitive endpoints to detect endocrine-related phenotypic effects,

including onset of puberty, uterine weight, and development of mammary tissue. This study will also assess maternal to fetal transfer of BPAF. The second aim is to undertake ADME studies to characterize the persistence of BPAF and changes in the metabolism of BPAF with life stage. The third aim is to conduct a multigenerational reproductive toxicity study and a developmental toxicity study. Additional endpoints may be added based on the knowledge gained from the first and second aims, the reported endocrine activities of BPAF and the reported effects of BPA.

b. Public Comments

Mr. Manuppello, PETA, said BPAF was nominated based on its moderate production and potential exposure of the general population, but no evidence was provided regarding exposure to BPAF. He said it is important that accurate exposure information be obtained before proceeding with the program.

Since BPAF is structurally similar to BPA, it seems likely that attempts to assess the reproductive and developmental toxicity will be subject to the same problems encountered with BPA, namely the reproducibility of endocrine effects at low doses. *In vivo* effects of BPAF were only observed after intraperitoneal injections or oral exposure to very high doses, which produced excessive toxicity. It is extremely unlikely that the proposed transgenerational study will produce clear evidence of toxicity at doses relevant to human exposure.

Although *in vitro* studies suggest that BPAF has endocrine activity, *in vivo* studies are inconclusive. Species differences in the metabolism of BPA may also confound the results with BPAF. PETA recommended that the potential for human exposure be accurately assessed prior to the development of the research program and that the metabolism of BPAF first be studied *in vitro* in human hepatocytes.

c. BSC Discussion

Dr. Mirsalis, a primary BSC discussant, said the rationale was clear and the proposed research program appropriate with the main focus being on the gestational and reproductive affects. The structural similarity between BPA and BFAF is adequate justification for the studies. He believed BPAF has a long half-life due to the fluorine atoms, which could result in environmental contamination even at low production levels and residues might be found in humans in the near future. The preliminary studies indicating that BPAF is a more potent estrogen than BPA is another reason to study it. He suggested that specific aim 2 be conducted before specific aim 1, so that more is known about BPAF's ADME before the large and expensive multigenerational studies are initiated. In addition, he suggested including standard genetic toxicology studies early in the program. Although, it is not clear there is public exposure, it is likely that BPAF is being discharged into the environment. Because of the environmental persistence of BPAF and the near hysteria among the public regarding BPA, he would elevate it to a high priority.

Dr. Vandenbergh, an *ad hoc* reviewer, agreed with Dr. Mirsalis' comments. He agreed that specific aim 2 should be conducted first to define the doses. BPAF should be examined for potential environmental contamination since BPAF binds to estrogen

receptor gamma, which is common in fish, but not found in mammals. He questioned why the research program is not studying BPAF for its possible effects on behavior and the central nervous system. He said evaluation of simple behavior endpoints such as sexually dimorphic behaviors would clarify whether BPAF has effects similar to those reported for BPA. He gave the study a moderate to high priority.

Dr. David Wegman asked about the uses of and exposure to BPAF as the only reported survey was over 25 years ago and he expected BPAF is used more now than in 1981. He asked if the NTP planned to work with NIOSH to update the use and exposure information. Dr. Toraason replied that NIOSH had been interested in updating the occupational exposure assessment (NOES) data for BPAF for at least 10 years. It is not easy for NIOSH to determine where a chemical is being used, since much of the information is proprietary. It might take 2-5 years to determine the exposure limits of BPAF despite it being a chemical of interest and the information important. Dr. Wegman said he is concerned about the apparent lack of collaboration between NTP and NIOSH in developing this kind of information. If NTP does not anticipate these needs and communicate with NIOSH, this information will never be collected to guide NTP in studying its nominations.

Dr. Kenneth Portier said the BPA studies were too small and lacked power to confirm the expected effects, resulting in uncertainty and an inability to make decisions. Many animals were wasted because the studies were too small. It is important that animal studies have statistical power especially when testing low-dose endocrine hypotheses, and it is essential that the data be analyzed correctly. Many of the factors that a good statistician would recommend, such as taking into account maternal effects and repeated measurements on the animals, were not included in the BPA studies and their analyses.

Dr. Bucher said the NTP has an interagency agreement (IAG) with NIOSH, which provides a formal mechanism to develop information on occupational exposures. As part of the IAG, NTP has funded some exposure assessment activities. Identifying where exposures are most likely occurring, getting access to industrial sites, and achieving cooperation from industry are usually difficult.

Dr. Stout said the NTP thought that obtaining a good measure of biological activity of BPAF through the transgenerational assay would be helpful, but would consider moving the ADME studies to specific aim 1. The NTP had discussed including BPA-specific endpoints in the initial studies; however, due to the limited information on production and use of BPAF and the limited toxicity data, the NTP thought a better approach would be to collect standard toxicity data first and then to look at endpoints based on the reported effects of BPA in the reproductive and developmental toxicity studies.

4. β-N- Methylamino-L-alanine

a. Presentation

Dr. Michael Sanders, NIEHS, presented the research concept for β-*N*-methylamino-L-alanine (L-BMAA), a non-protein amino acid. The NIEHS nomintated L-BMAA based on its potential for human exposure and evidence of neurotoxicity in animals. It is

produced by members of all five major taxonomic groups of Cyanobacteria (blue green algae), which are fairly ubiquitous in nature. L-BMAA has been detected in salt and fresh water, including drinking water sources, and in some plants and animals consumed as food. There are no known analyses for detecting L-BMAA in blue green algae dietary supplements, although, mycrocystins, which are potent hepatotoxins produced by cyanobacteria, have been detected in some of these products.

L-BMAA is toxic to motor neurons *in vitro* in the low micromolar range and causes neurotoxicity in animals at high doses. Following acute exposure, L-BMAA may form an active metabolite that reacts with glutamate receptors in motor neurons, causing damage and it can be incorporated into proteins of neural tissues damaging neuroproteins or serve as a reservoir for continuous low level activation of motor neurons.

Effects in non-human primates are similar to some symptoms observed in humans with amyotrophic lateral sclerosis, Parkinson's or Alzheimer's disease. Natives of Guam who consume L-BMAA in their cultural diet have an increased incidence of neurological disorders. L-BMAA has been detected in brain tissue of some Alzheimer's patients in North America; therefore, there is some concern for exposure to the general population.

The key issues are the extent of human exposure, and the uncertainty regarding L-BMAA's potency. There is a need to correlate internal dose with known exposure concentrations in an animal model and to investigate the nature of its interactions with proteins.

The goal of the research program is to characterize the toxicology of L-BMAA by conducting disposition and metabolism studies using radiolabeled chemical in rats and mice to study the extent and nature of its interaction with proteins, its persistence in tissues, and its elimination kinetics over time. Second, the biological activity of L-BMAA will be determined using *in vitro* techniques to compare its binding capacity with other glutamate-receptor-binding toxins. The third aim is to analyze dietary supplements for the presence of L-BMAA.

These studies should provide data to assess the proposed mechanism of toxicity of L-BMAA in humans, provide information about its biological activity, and determine its presence in dietary supplements. The data would be used to assess risk of exposure, provide public health guidance, and determine if further toxicity testing is needed.

b. BSC Discussion

Dr. Vandenbergh asked whether the quantity and number of algal blooms is increasing in the world. Dr. Sanders replied that the number of algal blooms is increasing, and thus the potential for exposure to L-BMAA.

Dr. Raymond Novak, the first BSC discussant, said the rationale and scope of the proposal was clearly presented and logically developed. The risk for human exposure and L-BMAA-mediated neurotoxicity is supported by existing literature and has the potential to be significant in the future. The proposed studies are important in view of the

potential for human exposure to L-BMAA in dietary supplements, food, and from algal growth in water systems shut down for significant amounts of time. These studies will provide valuable information on the toxicokinetics of L-BMAA and identify parameters critical to understanding the mechanisms by which this non-protein amino acid causes neurotoxicity.

He thought the determination of the ADME and PBPK following single and multiple oral doses of L-BMAA important as well as the *in vitro* studies that would be used to assess the biological activity of L-BMAA relative to known neurotoxicants. He thought the studies critical in view of an apparent latency period between the manifestation of acute effects, and the significant neurotoxic effects a decade or two later. He asked whether the *in vitro* studies would be performed in the presence and absence of bicarbonate. He wondered about protein catabolism and whether an adduct is formed that later causes serious disease. He suggested that certain factors such as fasting or other dietary alterations might impact normal ADME and encouraged the NTP to add these parameters to the studies as they may provide additional information on the metabolism and elimination of L-BMAA. He ranked the program as moderately high.

Dr. Michael Pino, the second BSC discussant, provided comment via telephone, as he was unable to attend the BSC meeting due to a medical condition. He said the rationale is well stated and based on existing data there is evidence that L-BMAA has neurotoxic properties that may be relevant to humans. He asked about annual sales of the blue green algae dietary supplements in the United States as a means of understanding the potential for human exposure. He agreed with Dr. Novak's comments about the scope and specific aims. He thought the ADME studies important considering the hypothesized latency of L-BMAA and agreed with measuring the amount of L-BMAA in dietary supplements to assess potential exposure of the general public. He thought the *in vitro* studies should be a low priority because there are data from *in vitro* studies and the chronic toxicology studies are warranted pending the outcomes of the other investigations. He gave the program a moderate priority.

Dr. Howard said the FDA cannot estimate the sales of blue-green algae dietary supplements and added that the toxicity of a specific dietary supplement would depend upon the type of algae used to make the supplement. There is no market survey of contamination and it is unknown whether some or all supplements are contaminated with L-BMAA.

Dr. Sanders said the *in vitro* assays could be designed so that the presence and absence of bicarbonate would be included. He thought the inclusion of various dietary regimens in the ADME studies would be useful. He said a 1999 reference estimated that over one million Americans and Canadians used blue-green algae supplements with a daily consumption up to 20 grams. He agreed that the *in vitro* experiments are important to more clearly define the potency issue.

5. Triclosan

a. Presentation

Dr. Howard, FDA, presented the research concept on triclosan. Triclosan is a bactericidal and bacteriostatic chemical used in a variety of personal care products including liquid soaps, dish detergents, deodorants, and dentrifice and oral rinses. Industrially it is used as a bactericide to control methycillin-resistant *Staphylococcus aureus* and other organisms. Triclosan is among the top seven wastewater-contaminants found in lakes and streams around the world. A private individual and the FDA nominated triclosan for dermal carcinogenicity testing because of its high level of topical use in humans, demonstrated transport through the mucosa and skin, and lack of dermal carcinogenicity data.

In humans, triclosan is absorbed via the oral and gastrointestinal mucosa, and through the skin at a rate of less than 10% with a half-life of 11 to 14 hours. It has been detected in human serum, urine and milk. The calculated average daily intake is 74 µg/kg/day. Mechanistically triclosan inhibits a bacterial type II fatty acid synthase, and intercalates into bacterial membranes reducing microbial viability. When heated at 600^{o}C in combination with textiles it is converted to 2,7 or 2,8- dichlorodibenzo-*p*-dioxin. If hypochlorite is added to the combustion mixture, tri and tetrachlorodibenzo-*p*-dioxin is formed although the efficiency of this process is unknown. In the environment triclosan is photodehalogenated and photocleaved to phenols but it can also undergo photo-induced ring closure to form dichlorodibenzo-*p*-dioxins. Triclosan is not mutagenic and has low acute and subchronic oral toxicity in rodents, rabbits and dogs. It is an estrogen antagonist in frogs and elicits estrogenic effects on MCF-7 cells in culture.

Since 1978 the FDA has classified triclosan as a category 3 product because there are insufficient safety data for a complete toxicological assessment, primarily due to a lack of dermal carcinogenicity data. The FDA requests these studies because of the significant level of exposure to triclosan from a variety of products throughout all life stages. The primary aim is to conduct a dermal carcinogenicity study of topically applied triclosan. A secondary aim is to determine if dichorodibenzo-*p*-dioxins are formed *in vivo* and/or *in vitro*. The pharmacokinetics of triclosan in mice will be determined following dermal application to establish a dose range that can be used in a dermal carcinogenicity study. The possible formation of phenols or dichlorodibenzo-*p*-dioxins on the skin following exposure to light will also be determined. If these compounds are significantly formed, photocarcinogenicity studies may be warranted.

b. Public Comments

Mr. Manuppello, PETA, said PETA does not believe a dermal bioassay is needed as triclosan has been used in a wide variety of consumer products for over 40 years without evidence of skin irritation. In 2001, CIBA submitted a position paper addressing the need for a dermal carcinogenicity study following the completion of a long-term hamster study in response to a FDA request. All the test vehicles were too harsh for a long-term study. Key findings were that triclosan does not have the profile or biological activity of any known skin carcinogen, it is nongenotoxic, and it does not cause hyperproliferative changes in the skin with typical use. NTP agreed that triclosan is not genotoxic, has low activity in acute studies, and found no maternal or fetal toxicity in rodents and rabbits up to the highest doses tested in a battery of reproductive tests. The only concern noted is

dermal irritation reported in a 1998 Colgate Palmolive subchronic study in rats. The FDA has taken no further action on the assessment since 2001 despite the submission of 27 studies in 2003. It is imperative that FDA reconsider its call for a dermal carcinogenicity study because these studies will consume hundreds of mice and are likely problematic considering previous difficulties in identifying appropriate vehicles. If the preliminary phase of the proposed NTP studies to determine dermal penetration and steady-state levels in the skin of mice and the kinetics of triclosan's photodecomposition are necessary, PETA urges the FDA to consider using excised skin as outlined in OECD 421 or a reconstituted human skin model.

c. BSC Discussion

Dr. Russell Cattley, the first BSC discussant, said the rationale for this research concept was presented clearly. Although there is widespread use of triclosan as an antimicrobial agent, there does not appear to be a robust evaluation of its carcinogenic hazard following dermal exposure, and there is uncertainty about the validity of studies in which rats and hamsters were exposed via the oral route. There is evidence of cutaneous absorption of triclosan by mouse skin, but because of it thinness he was uncertain about the feasibility of evaluating the uptake of triclosan in mouse skin. Rather, he suggested that *in vitro* approaches be considered to evaluate toxicity at the cellular or molecular level and to assess metabolism in skin cells. If *in vitro* assays can be developed, then the possibility of species extrapolation experiments and assessment of triclosan's photodecomposition to dioxins would be possible. If the *in vitro* studies provide evidence for the formation of dioxins, the priority of the proposal would increase.

The proposed one-month and chronic studies are traditional approaches for assessing dermal carcinogenicity. Care will be needed with the dermal studies to prevent simultaneous oral exposure to triclosan. He supported postponing the photocarcinogenicity study until the outcome of the one-month study is known. He referred to several recent references relating to potential endocrine effects including interference with testosterone metabolism. He suggested that the NTP test whether triclosan interferes with testosterone metabolism in *in vitro* assays and if so, it may be necessary to test it in *in vivo* studies.

Dr. Vandenbergh, an *ad hoc* reviewer, was concerned about the findings in three recent publications that showed that triclosan binds to the estrogen, androgen, and possibly thyroid hormone receptors in rodent cells. One study indicated that binding occurs at extremely low levels. There were indications that triclosan also may be active through mechanisms other than binding to the estrogen receptor. He was not sure whether NTP should wait until the literature is clearer before adding studies to investigate reproductive outcomes.

He was also concerned because triclosan is found in relatively high concentrations as compared to other contaminants in the environment. Since the breakdown of sewage depends on microbial action, he questioned whether triclosan as an antibacterial agent could affect the microbial biota. He gave the study a moderate priority because the FDA needs the data and said including studies to assess endocrine effects might prove

informative.

Dr. Jim Riviere, the second BSC discussant, said dermal carcinogenicity studies are crucial due to the wide use of triclosan and the potential for its photodegradation to dioxins. He questioned the use of mice for the dermal studies because penetration of substances to the basal cells of mouse skin is rapid and not representative of absorption in other species. He suggested that a species with more complex skin morphology be used in the studies. Although UV-induced skin cancer has increased, there is little or no understanding whether triclosan forms dioxins in the presence of light, and he encouraged the NTP to address this important question.

Dr. Kerkvliet asked Dr. Howard to comment on the relevance of the possible formation of 2,7 dichlorodibenzo-*p*-dioxin in terms of skin cancer. Dr. Howard replied that it is unknown whether 2,7 dichlorodibenzo-*p*-dioxin is formed in the skin, but when triclosan is dispersed in water and irradiated with a strong source of UV light, it is formed. Triclosan may preferentially bioconcentrate in the stratum corneum or upper layers of the epidermis where photons have a better chance to interact and convert it to dioxins. Formation of 2,7 dichlorodibenzo-*p*-dioxin could be addressed easily in *in vitro* studies. He agreed with Dr. Riviere about the structure of the skin and that human-relevant models should be used for the *in vitro* studies. He doubted whether 2,7 dichlorodibenzo-*p*-dioxin would have similar tumor-promoting properties as 2,3,7,8, tetrachlorodibenzo-*p*-dioxin (TCDD) because it is many orders of magnitude less potent than TCDD, but the formation of 2,7 dichlorodibenzo-*p*-dioxin should be ruled out as a confounder of any dermal effects.

Dr. Novak wondered if triclosan would function as a promoter if it were applied after UV irradiation damaged the skin. Dr. Howard said the dermal carcinogenicity studies would not involve UV exposure. However, he had not considered that triclosan might be a promoter and said a study to test this hypothesis would be considered.

Dr. Kenny Crump did not believe a dermal carcinogenicity study necessary because of the data available on oral carcinogenicity. He suggested that the pharmacokinetic data and predicted doses in target tissue be evaluated first to assess whether dermal exposure would provide additional information.

Dr. Howard thanked the BSC and the public for their helpful comments. He said the FDA had reviewed the oral carcinogenicity data and the pharmacokinetic profile carefully. He did not believe that the skin-carcinogenicity hazard of triclosan could be modeled from the oral data because it had a different pharmacokinetic profile with dermal absorption. He said the principal focus for the FDA is dermal carcinogenicity, but he also would consider studying triclosan's reproductive toxicity and potential for endocrine disruption.

6. Hydroxyurea

a. Presentation

Dr. Masten, NIEHS, provided background information on hydroxyurea, an antimetabolite

that inhibits ribonucleotide reductase resulting in inhibition of DNA synthesis, cell cycle arrest, and cytotoxicity. When these effects occur in red blood cells, the production of fetal hemoglobin increases, which is the basis for treating sickle cell disease (SCD). Hydroxyurea was approved by the FDA for treatment of SCD in 1998 but has been used since 1967 to treat certain neoplastic diseases. There are off-label uses to treat myoproliferative diseases, thalassemia, and HIV infection. It is the only treatment for sickle cell disease used in children aside from blood transfusion.

A private individual nominated hydroxyurea in 2006 for carcinogenicity studies, due to its widespread use in the treatment of SCD and other myeloproliferative diseases, demonstrated mutagenicity, and safety concerns associated with long-term use. The NTP deferred review of the nomination at that time because of an ongoing Center for the Evaluation of Risks to Human Reproduction (CERHR) evaluation. An expert panel convened by CERHR identified critical data needs including multigenerational studies to assess long-term effects of pre- and postnatal exposures of hydroxyurea on developmental neurotoxicity, reproductive function, and carcinogenicity. There is some evidence of reproductive toxicity in case reports of sickle cell patients receiving hydroxyurea where low sperm count and reduced sperm motility have been observed. Multiple effects on development and reproduction have been reported in rodents. Blood concentrations of hydroxyurea in animal studies where toxicities were observed were similar to those in patients on therapy.

Considerable data indicates hydroxyurea is genotoxic. Case reports and small cohort studies show a possible association with acute leukemia and certain skin cancers. In an NCI/NTP carcinogenicity study reported in 1977 there was an increased incidence of mammary tumors following intraperitoneal injection of hydroxyurea for six months. IARC considered the evidence for carcinogenicity inadequate and classified hydroxyurea as a non-classifiable Group 3 carcinogen.

There are no long-term experimental animal studies or rigorous toxicology studies to address the safety issues regarding chronic exposure to hydroxyurea beginning early in life. Despite these data inadequacies, in a published statement from an NIH Consensus Development Conference the risks associated with hydroxyurea were considered acceptable compared to the risks of untreating sickle cell disease. Attendees identified the need for further studies to better understand adverse developmental and reproductive effects and carcinogenic risks, although it is uncertain if additional animal toxicology studies would alter current clinical treatment guidelines.

The NTP proposes not to develop a research program for hydroxyurea at this time. There are a number of ongoing clinical trials as well as follow-up analyses from prior clinical trials to address safety endpoints such as reproductive toxicity in men. NTP hopes the research needs identified by CERHR, the Agency of Health Care Research and Quality, and the NIH consensus development conference, will be addressed though new human studies. The NTP will continue to monitor the progress of hydroxyurea safety research and revisit the need for additional animal studies in the future.

b. BSC Discussion

Dr. Cattley asked for the dose-limiting toxicity for therapeutic use, and Dr. Masten replied that it was primarily hematotoxicity when used for SCD. There is considerable trial and error in the titration of an appropriate dose for therapy.

Dr. Friedman-Jiménez said long-term prospective carcinogenicity studies in humans would not provide useful information in the near future. He asked if there are any funded retrospective epidemiological studies underway to address the human carcinogenicity issue, and Dr. Masten said he was not aware of any such studies. Dr. Friedman-Jiménez then asked what the plan is to answer the data gap on human carcinogenicity. Dr. Masten replied that only an animal study would address that gap in the relative near-term, but NTP is not proposing such a study at this time.

Dr. Howard said Dr. Friedman Jimenez's call for a retrospective analysis is a principle recommendation from the CERHR panel, which suggested setting up an active registry, since the population at risk is small and easily identifiable. The drug has the strongest warning label mandated by the FDA, which clearly states the hazards associated with its use including mutagenicity, clastogenicity, its potential to cause fetal harm, and its likelihood of being a human carcinogen. This information is not necessarily conveyed to the patient, but the intent is for the physician to know the relative risk of prescribing hydroxyurea. The FDA supports the proposal for an epidemiology study to define human risk. In view of the significant off-label use and potential serious health impact to pediatric populations, the FDA does not object to animal studies for understanding any risk associated with hydroxyurea, especially if reproductive and developmental endpoints were included.

Dr. Carney said the resolution of the issues is more in the range of a human clinical study with perhaps some animal work. Hydroxyurea has been shown to be teratogenic in eight different species and thus there is no value in further animal studies. Recent studies confirmed that hydroxyurea affects sperm motility, sperm counts, and sperm morphology in men. He did not believe that a reproductive study in animals would add any new scientific data, but felt there was some merit to long-term studies to address carcinogenicity.

Dr. Portier agreed that there did not appear to be an added benefit for conducting additional animal studies and the need for human research is questionable, as a highly toxic material is being used for people with an extreme health condition. He gave the concept a low priority.

Dr. Friedman-Jiménez asked whether the National Heart, Lung and Blood Institute (NHLBI) had funded any research on alternative treatments for sickle cell anemia because of its presumed carcinogenicity. Although SCD is a debilitating disease, some patients may refuse treatment because of its carcinogenic potential. Dr. Masten replied that gene therapy is being pursued, and two compounds, namely clotrimazole, an antifungal agent, and decitabine, an anti-metabolite with serious reproductive and developmental risks, are being evaluated in clinical trials. Sickle cell is a difficult disease

to treat; any therapy has a spectrum of risks that have to be considered in light of the benefit.

Dr. Friedman-Jiménez made an analogy to tamoxifen that was developed and used for the prevention of breast cancer, even though studies found it to be a carcinogen. This latter finding prompted the development of raloxifene, which is also a preventative that does not cause cancer. He wondered if additional data on carcinogenicity might upgrade the IARC classification of hydroxyurea, which in turn would drive the development of alternative treatments. From a clinical perspective it is very important to have an effective and less toxic treatment for sickle cell anemia.

Dr. Wegman said he was troubled by the summary of risks from the NIH consensus development conference that considered the treatment acceptable compared to the sequelae of the disease. Communication is important, and the fact that the label states it is a carcinogen may confuse patients as to whether they should take the medication or not. He worried that the carcinogenic statement is based on an inadequate study and was skeptical about assembling a cohort to obtain carcinogenic information. There might be more of an impetus to develop a drug if it were better known what the risk of cancer is in this population.

Dr. Mirsalis agreed that an epidemiology study is unlikely to provide useful information because these patients may not live long enough to succumb to cancer. However, if hydroxyurea prolongs their life, this in itself will increase their risk for cancer because age is a risk factor for developing cancer. He believed that if hydroxyurea were found to be a rodent carcinogen, it would not encourage the development of new drugs for the disease, nor would it change patients' therapy, risk, or whether patients would take the drug. Of the 2000 or so drugs on the market, 50% are rodent carcinogens. He supported the recommendation for no studies.

Dr. Crump said he supported the recommendation not to go forward with an animal study. He believed that more definitive information would be gleaned from a retrospective epidemiological study; however, it is unclear how such data would affect hydroxyurea treatment or development of alternative treatments. He thought the discussion on tamoxifen was interesting, but he wondered if the search for new therapies was based on its carcinogenicity in an animal study or human epidemiology data.

Dr. Friedman-Jiménez confirmed that the data were based on both human and animal studies. He asked whether there are alternative plans to address this public health and clinical issue if the BSC does not recommend animal studies. He believed the BSC could contribute meaningfully to addressing data needs for hydroxyurea, but was choosing not to do so.

Dr. Kerkvliet asked the BSC if there were any chance an animal study would show that hydroxyurea was not carcinogenic. Dr. Mirsalis replied that although one would presume methotrexate to be carcinogenic, it is not. Both hydroxyurea and methotrexate interfere with DNA replication. He would be surprised if hydroxyurea were non-carcinogenic in

rodents.

V. Criteria for Evaluation of Outcomes in Reproductive, Developmental, and Immunotoxicology Studies

1. Background to Criteria for NTP Non-Cancer Studies

a. Presentation

Dr. Paul Foster, NIEHS/NTP, provided background for development of the criteria for evaluating outcomes in reproductive, developmental, and immunotoxicology studies. He explained that NTP's goal is to employ the same rigorous standards used historically to review carcinogenicity bioassays to review NTP "non-cancer" studies. Efforts thus far to enhance the rigor of these studies include training workshops for NTP and contractor pathologists, establishment of Pathology Working Groups to review and agree on the diagnosis of critical lesions in NTP non-cancer studies, and peer review by the NTP BSC Technical Reports Review Subcommittee of the draft reports for multigenerational studies. NTP desires to have consistent criteria to evaluate NTP study outcomes.

Dr. Foster explained that the NTP has long employed specific conclusion statements, approved by the BSC, for its "toxicology and carcinogenesis" studies that present a "level of evidence" with regard to the carcinogenic potential of the test substance for each sex/species within an individual study. These levels are: clear evidence, some evidence, equivocal evidence, no evidence, and inadequate study. This approach provides uniformity in the reporting of NTP findings across different studies on the same substance and across studies on different test substances. He gave an example of conclusion statements for the carcinogenic activity of sodium dichromate. The NTP realized that "levels of evidence" criteria were needed for its non-cancer studies to ensure comparability in the reporting of outcomes across studies, i.e., developing similar criteria for NTP non-cancer studies would allow for "level of evidence" comparisons to ensure the findings are evaluated similarly for different test substances and for different study types for the same test substance. NTP discipline leaders in immunotoxicology and reproductive and developmental toxicology developed draft criteria for each study type and some guidance for applying the criteria.

He presented some issues that the NTP considered in developing the draft criteria including (1) conclusions statements are hazard-based, not risk-based, to facilitate comparison across test substances for the same study types; (2) many of NTP's non-cancer toxicity studies include multiple (inter-related) endpoints, which is different from the cancer studies; (3) applying the NTP cancer study "levels of evidence" approach to non-cancer studies would require some "finessing" to achieve the desired level of consistency; (4) it is desirable to use a graded (hazard identification) "level of evidence" scheme for expressing conclusions; (5) endpoints that affect overall system function merit the highest level of evidence ("clear evidence" of toxicity); and (6) applying the criteria to clearly positive or negative results should be straightforward, whereas findings at the boundaries would present more difficulty. The NTP staff then took steps to refine

the draft criteria including "in house" exercises, informal sharing with external colleagues for feedback, and the formation of BSC working groups to provide input.
The NTP convened two BSC working groups, one with expertise in immunotoxicology and a second with expertise in reproductive and developmental toxicology. The two working groups met separately to discuss the proposed NTP criteria, assess their utility and suitability and make recommendations on any needed changes. Each working group was provided the draft criteria, guidance on their application, and case studies to explore the utility of the criteria.

Dr. Foster then discussed some implications inherent in the adoption of these criteria: (1) more consistency in the evaluation of the non-cancer studies, (2) the potential for the studies to be noted as "authoritative" by certain regulatory bodies like the cancer studies, (3) the need for appropriate expertise on the BSC (or BSC subcommittees) to review the studies, and (4) the potential adoption of the criteria by other groups.

b. BSC Discussion

Dr. Cattley asked if the study design is well defined at the start with every compound. Dr. Foster said there is a basic minimum study design, which is adapted or amended (e.g., with additional endpoints) to address the specific goals of the study.

2. Levels of Evidence Criteria for NTP Immunotoxicology Studies

a. Presentation

Dr. Kerkvliet, chair of the NTP BSC Immunotoxicology Criteria Working Group (ICWG), presented the working group report. Dr. Portier served as BSC chair during the presentation and discussion. He explained that the BSC would be voting on whether to accept the report. Dr. Kerkvliet said the meeting was held August 13 – 14, 2008, in Arlington, VA and the ICWG was comprised of people from academia, industry, and regulatory groups with a broad range of expertise in immunotoxicology and hazard assessment. Expert technical advisors were also present at the meeting. The charge to the working group was, "*Evaluate the suitability and utility of the proposed criteria for describing the results from individual NTP immunotoxicology studies to indicate the strength of the evidence for their conclusions."* The NTP provided the ICWG with draft "levels of evidence" criteria, which they applied to 30 case studies. The case studies were typical of the different types of data that could be encountered in immunotoxicology studies. She explained some of the issues considered in applying the draft criteria such as biological plausibility, consistency of dose-response, and functional versus non-functional changes. The ICWG reviewed the case studies individually, discussed the results as a group, proposed revisions to the draft criteria, created some key points for consideration in applying the criteria, and prepared their report.

Dr. Kerkvliet then explained the levels of evidence the ICWG drafted for evaluating immune system toxicity:

Clear Evidence of Toxicity to the Immune System
- Is demonstrated by data that indicate a clear treatment-related (considering the magnitude and the dose-response) effect on more than one functional parameter and/or a disease resistance assay that is not a secondary effect of overt systemic toxicity, or
- Is demonstrated by data that indicate treatment-related effects on one functional assay and additional endpoints that indicate biological plausibility.

Some Evidence of Toxicity to the Immune System
- Is demonstrated by data that indicate a treatment-related effect on one functional parameter with no other supporting data, or
- Is demonstrated by data that indicate treatment-related changes in multiple non-functional parameters without robust changes in a functional immune parameter or a disease resistance assay, or
- Is demonstrated by data that indicate non-dose-related effects on functional parameters or a disease resistance assay with other data providing biological plausibility.

Equivocal Evidence of Toxicity to the Immune System
- Is demonstrated by data that indicate non-dose-related effects on functional parameters or a disease resistance assay without other data providing biological plausibility, or
- Is demonstrated by data that indicate treatment-related changes in a single non-functional parameter without changes in a functional immune parameter or a disease resistance assay, or
- Is demonstrated by data that indicate immune effects at dose(s) that produce evidence of overt systemic toxicity, or
- Is demonstrated by data that are conflicting in repeat studies.

No Evidence of Toxicity to the Immune System
- Is demonstrated by data from studies with appropriate experimental design and conduct that indicate no evidence of biologically relevant changes in immune parameters.

Dr. Kerkvliet said the ICWG did not include an "insufficient evidence" level because of the understanding that any published NTP study would have sufficient evidence. She then reviewed some of the key points discussed by the ICWG, including (1) immune response enhancement or suppression by toxicants, (2) the impact of overt toxicity, (3) the intended pharmacology of the chemical, (4) transient immune effects, (5) consideration of biological plausibility, and (6) the purpose of the criteria (for hazard identification only, not risk assessment). She concluded by stating that the ICWG meeting was a great success.

b. BSC Discussion

Mr. Janzen asked about non-dose-related effects. Dr. Kerkvliet explained that it is sometimes just the nature of immune responses, e.g., it is not uncommon to have an

antibody response that is enhanced at only the lowest dose but suppressed at higher doses. Dr. Germolec added that it could depend on the target cell population, e.g., a low-dose effect on regulatory cells and a different treatment-related effect at another dose, that can look like conflicting data, unless the mechanism is understood. Also, sometimes with a single parameter, there is no dose-response because everything may be suppressed or enhanced.

Dr. Crump asked for clarification about the terminology used to describe "non-dose-related." He considered "dose-related" and "treatment-related" to mean the same thing, and suggested using "non-monotonic dose-response" instead. Dr. Wegman was not sure who would need to understand how to apply the criteria and, having applied them, what they mean. Dr. Kerkvliet explained that the criteria are for use by NTP, but the ICWG realized that the criteria were likely to be used more broadly. The criteria were not written for the general public. Dr. Wegman said, as an epidemiologist, he was confused by the variation in how levels of evidence are assessed, but that the criteria might work for a well-informed toxicology community. He agreed with Dr. Crump regarding the issue of nonlinear dose-response and suggested that the criteria include a discussion of statistical power. He stated that carcinogenic classification is much easier, whereas creating these criteria is a harder task due to having to move beyond a dichotomous classification to a more complex system. He considered the criteria a great start, but not yet complete.

Dr. Wegman said if the audience is immunotoxicologists, the criteria are useful, but beyond them, there could be difficulty in their communication. Dr. Foster explained that the criteria are for drawing conclusions on NTP studies only. The NTP is attempting to compare across NTP studies, using a fairly standard protocol that is adequately powered for an assessment of toxicity and permits comparisons across chemicals. The conclusion statements will be hazard-based. Dr. Wegman said cross-study comparisons used within NTP are fine, but when the information is published in the technical report, the greater community should understand the terms as well. He said the carcinogenesis criteria are fairly well understood by a wide audience. Dr. Bucher said, initially, the cancer criteria were not well understood and acknowledged that initially neither will the immunotoxicology criteria be well understood. He agreed that criteria for immunotoxicology studies are a more complicated issue because of the dynamic and self-correcting nature of the immune system and the wide variety of endpoints that have to be interpreted by experienced immunotoxicologists. The goal for this categorization system is a consistent approach toward assigning a level of concern for a spectrum of findings. It will take some time before the regulatory community can become comfortable with and understand the criteria.

Dr. Kerkvliet said the EPA members of the ICWG were enthusiastic about the criteria. Dr. Wegman said the toxicology community would not allow the epidemiology community to get away with this amount of ambiguity. Dr. Howard said the criteria would have great utility and help alert the regulatory agency about a potential hazard, but the agency would not act just based upon the NTP's conclusion statement, but would take the data, reconstruct the risk profile, and then make a risk management decision. Dr.

Friedman-Jiménez said the criteria are to be used as an internal system for grading individual studies, not for integrating information from several studies or for supporting regulatory decisions. Dr. Foster reiterated that NTP studies are hazard-based assessments; the NTP usually does not have exposure data. It is an assessment of the strength of evidence from a particular study that would enable the reader to draw conclusions about the relative weight of evidence for toxicity; it is not a risk assessment.

Dr. Cattley, a lead discussant, asked if the NTP would be looking at more than one study when applying the criteria. Dr. Foster explained that NTP does not repeat multigenerational studies, but will use two species in a developmental study for the same compound, which often give slightly different information. Dr. Cattley asked about using two different functional assays to obtain one conclusion about the test compound. Dr. Germolec said if a positive result were found in a single parameter, the test would be repeated. The purpose of the criteria is to look at the sum total of the data in the study. NTP range-finding studies assess the humoral, cell-mediated, and innate immune responses and all of those data would be used. NTP would not do a host resistance assay, which is considered a tier 2 study, unless a positive result was seen in a basic immune test. The NTP selects the appropriate studies based on where a deficit is seen in immune function. NTP's conclusion is made based on the results of a number of individual assays.

Dr. Bunton asked about the rationale for conducting a test, such as whether a signal from another study or a suspicion that the chemical might be an immunotoxicant. Dr. Germolec said there are multiple sources driving the testing. Sometimes compounds are nominated due to clinical information that, perhaps, there is some unintended suppression by a therapeutic. Other NTP studies may indicate a potential for immuntoxicity, e.g., 14-day studies showing changes in hematological parameters, or data from histopathology studies, or studies showing changes in organ weights. Dr. Bunton asked about controlling for stress when doing a range-finding study. Dr. Germolec said the tests are done at dose levels that do not produce systemic toxicity.

Dr. Friedman-Jiménez asked about the integration of studies including the use of human clinical information. Dr. Germolec replied that the criteria were developed specifically to interpret data from NTP studies, though information from other studies about potential immune effects may act as a trigger for the NTP studies. The criteria are simply to indicate the immuntoxicologic potential of a substance, based on the data from the tests the NTP conducted. Dr. Bucher said the first words of the conclusion statements are, "Under the conditions of this study..." Dr. Friedman-Jiménez asked about multiple studies done under different conditions. Dr. Bucher said there would then be more than one conclusion. Dr. Portier said one study could have multiple experiments and protocols; another study done for another reason could provide conflicting conclusions. Dr. Mirsalis said conflicting conclusions have occurred with the carcinogenicity studies wherein sometimes there is clear evidence in male rats and no evidence in female mice.

Dr. Novak, a lead discussant, said he concurred with much of the discussion that had transpired. He agreed with the definition of "no evidence" of toxicity to the immune

system, but the rest is confusing. He asked about the use of the terms nonfunctional and functional, and suggested using "observational" in lieu of "nonfunctional" in the criteria. He saw no need for including interjected statements in the criteria that detract from the definitions. He suggested wordsmithing the criteria to clarify the concepts. Dr. Kerkvliet said that she should have presented the tier of assays used by NTP that would have put the criteria in context. Dr. Novak said the assays are described in the preamble of the report, which was helpful. He suggested being careful about the choice of words when spanning different audiences; to him a "nonfunctional" assay is an assay that did not work.

Dr. Cattley said he would like the report to include more information about the case studies. Most assays can detect substances that cause immunosuppression, but there are fewer assays that can detect immune stimulation or autoimmunity. Dr. Germolec said the case studies were a mix of NTP studies and other published data, some of which the ICWG submitted. Their goal was to explore the boundaries, sometimes using incomplete data sets, and the issue of intended versus unintended suppression. The case studies were heavily weighted to NTP chemicals (~ 20 of the 30 chemicals) and represented a spectrum of data. The non-NTP case studies were meant to stimulate discussion among the ICWG and to try and identify when the draft criteria did and did not work. Dr. Cattley suggested a level for "inadequate" studies. If the only evidence for immune system toxicity occurs at considerable systemic toxicity, he would consider that "no evidence" or "inadequate" rather than "equivocal."

Dr. Cattley asked about disease resistance assays, which he thought should not be as heavily weighted as functional assays because of the possibility of nonspecific effects. Dr. Germolec said the immuntoxicology community views host resistance assays as the "gold standard" and gives them the most weight. The NTP uses bacterial challenge models that operate by the same mechanism in humans and rodents and does not use death as an endpoint, but instead uses sensitive parameters, such as colony or tumor counts. The NTP looks at whether the chemical itself could potentially induce toxicity or cause a decreased number of organisms. The test includes multiple doses of the chemical and multiple doses of the organism. Dr. Germolec said the disease resistance assays, which are second tier, are not done unless the NTP has some information about the mechanism(s) of toxicity. Because the immune system is somewhat redundant, an effect with a single endpoint may not translate into changes in disease resistance. The disease resistance assays provide concrete evidence that the immune system, as a whole, and not just a single parameter, is affected. Dr. Kerkvliet added that the host resistance assays are used for validating immunotoxicity and that there are instances in which changes in host resistance do not translate into measurable effects in immune responses.

Dr. Riviere asked what the term "dose-related" means; whether she was referring to a graded dose response or that only one out of four doses has an effect. Dr. Kerkvliet indicated that dose-related means the same as dose-dependent, which is not necessarily monotonic.

Dr. Howard applauded the ICWG's effort. He said the extreme rigor of the pathology evaluations for NTP carcinogenicity studies creates tremendous confidence within the regulatory community and scientific community, in general, regarding the data reported in NTP Technical reports. He asked if the same rigor exists for the functional immunotoxicology assays. Dr. Germolec said there have been a number of interlaboratory validation studies of both the functional tests and the host resistance assays. There have also been efforts to look at how predictive the individual functional tests are. The immune assays that tend to be the most sensitive and predictive are the ones that integrate multiple types of immune cells or multiple immune endpoints; e.g., the antibody-forming cell assay is the most sensitive and predictive functional test that is not a disease resistance assay. NTP always includes a number of different positive controls, specific for the immune tests, and has determined which positive controls work best. Dr. Howard said some of the tests have OECD guidelines.

Dr. Friedman-Jiménez asked about the possibility of producing a report similar to the NTP Report on Carcinogens, that integrates multiple studies to come up with one judgment on a compound. Dr. Friedman-Jiménez said the NTP could create a report on immunotoxicology that would integrate animal, mechanistic, epidemiologic, and clinical evidence to make a judgment whether a substance is immunotoxic. Dr. Bucher replied that this suggestion would be an interesting recommendation for the BSC to make. Dr. Kerkvliet said the NTP only undertakes immunotoxicology studies if there is preliminary evidence to suggest that a substance affects the immune system, so she was unsure if the database would be robust.

Dr. Crump said the criteria do not mention statistical significance, which is also true of the cancer criteria, though statistical significance is given a lot of consideration when making a call on carcinogenicity. He thought "treatment-related" implied statistical significance and it should be added to the guidelines. He said the idea of statistical power is important, especially in the "no evidence" category. The criteria state, "appropriate experimental design," which should include the concept of power. He suggested not using the term "non-dose-related," because it implies non-treatment-related. He said the terminology, e.g., the use of "observational" versus "nonfunctional," should be as general as possible so individuals not familiar with the immunotoxicology field will understand. Dr. Novak said defining the tests in the preamble as "functional" or "observational" would be helpful in clarifying the terminology.

Dr. Toraason encouraged including the "inadequate study" level in the criteria. He also expressed concern about using the terms "enhancement, stimulation, robust, and significant."

Dr. Portier asked the BSC to address whether the report is acceptable with the modifications suggested. Dr. Wegman thought the report is good start and that it would evolve. He suggested amending to title to include "Version 1," to suggest a "work in progress." Dr. Portier said one suggestion is to see the report come back to BSC with Version 2. Option 2 is to approve tentatively, with the recommended changes, and with the BSC seeing the changes before giving complete approval.

Dr. Bucher said the discussion had been great with excellent suggestions from the BSC. The report of the ICWG is advice to the NTP for consideration when making final adjustments to the criteria and putting them into practice. He said it is not necessary to attain perfection before voting to accept the report and that all of the collected comments would be taken into consideration. Mr. Janzen asked about the changes that had been discussed and was assured that the comments had been captured and would be considered.

Mr. Janzen applauded the ICWG for addressing the topic and for their progress to date. Dr. Bucher said the vote would be to accept the report of the ICWG and the comments made today as advice to NTP. When NTP finalizes the criteria, they would be brought to the BSC. Dr. Mirsalis moved that the BSC accept the report, and that a transcript of the comments be included as an appendix to the report to show the discussion and the issues of concern. Mr. Janzen seconded the motion. Dr. Friedman-Jiménez thought it was important to move forward and that the experience gained by using the criteria would be helpful in coming up with Version 2. Dr. Crump said implementation of the critieria should be delayed until they are in final form because studies evaluated under early guidelines cannot be compared with studies evaluated under amended guidelines. Dr. Kerkvliet agreed and said the preamble could easily be amended without amending the levels of evidence. Dr. Portier said he understood from the discussion that use of the five levels is appropriate, but that additional description, including a preamble, and clarification of language are needed to make the criteria more functional. The BSC agreed with Dr. Portier's summary. The motion passed unanimously with 14 yes votes, 0 no votes and 0 abstentions.

3. Levels of Evidence Criteria for NTP Reproductive Toxicology Studies

a. Presentation

Dr. Carney said the BSC Reproductive and Developmental Criteria Working Group (RDCWG) addressed reproductive and developmental toxicology studies. The charges to the working group were, *"To evaluate the suitability and utility of proposed criteria for describing the results from individual NTP reproductive toxicology studies to indicate the strength of the evidence for their conclusions,"* and, *"To evaluate the suitability and utility of proposed criteria for describing the results from individual NTP developmental toxicology studies to indicate the strength of the evidence for their conclusions."* He explained that NTP studies are extremely large and with sufficient statistical power.

Dr. Carney explained that the process for development of the reproductive studies criteria was very similar to the process described for the immuntoxicology criteria. Strawman "levels of evidence" criteria were supplied to the working group along with study examples that the RDCWG scored individually. The working group then reviewed the results, made adjustments to the criteria, and developed key issues to be used in applying the criteria. He said the working group included reproductive toxicologists and risk assessors from government, industry, and academia. He explained that the NTP describes the results of individual studies of chemical agents, and notes the strength of the evidence for conclusions regarding each study. Negative results do not necessarily imply that a

chemical is not a reproductive toxicant, but only that the chemical is not a reproductive toxicant under the specific conditions of that study. Positive results are assumed to be relevant to humans, unless data are available which demonstrate otherwise. Given that developmental events are intertwined in the reproductive process, effects on developmental toxicity may be detected in reproductive studies. The "levels of evidence" statements describe only reproductive hazard. The determination of risk to humans requires exposure data that are not considered in these summary statements. Five categories of evidence are included in the criteria: clear, some, equivocal, no, and inadequate study.

The RDCWG recommended that a study's lowest observed adverse effect level (LOAEL) be reported for positive results, and that the highest dose level (NOAEL) tested be reported for the "no evidence" category. The RDCWG considered the dose as a fundamental property of the compound. Application of these criteria requires professional judgment by individuals with working knowledge of the studies, which is a potential challenge for implementing the criteria.

Dr. Carney then explained the levels of evidence the working group drafted for evaluating reproductive toxicity:

Clear Evidence of Reproductive Toxicity

- Demonstrated by a dose-related[1] effect on fertility or fecundity, or by changes in multiple interrelated reproductive parameters of sufficient magnitude that by weight of evidence implies a compromise in reproductive function. A statement to the effect of "This study has a lowest observed adverse effect level of XXXX mg/kg/d for reproductive toxicity" should accompany the evidence statement.

[1]The term "dose-related" describes any dose relationship, recognizing that the treatment-related responses for some endpoints may be non-monotonic due to saturation of exposure or effect, overlapping dose-response behaviors, change in manifestation of the effect at different dose levels, or other phenomena.

Some Evidence of Reproductive Toxicity

- Demonstrated by deficits in reproductive parameters, the net impact of which is judged by weight of evidence to have potential to compromise reproductive function. Relative to clear evidence of reproductive toxicity, such effects would be characterized by greater uncertainties or weaker relationships with regard to dose, severity, magnitude, incidence, persistence and/or decreased concordance among affected endpoints.
- A statement to the effect of "This study has a lowest observed adverse effect level of XXXX mg/kg/d for reproductive toxicity" should accompany the evidence statement, except in those instances in which the "some" classification has been based on uncertainties about the dose relationship that precludes confident determination of the LOAEL.

Equivocal Evidence of Reproductive Toxicity

- Demonstrated by marginal or discordant deficits in reproductive parameters that may or may not be related to the test article.

No Evidence of Reproductive Toxicity

- Demonstrated by data from a well-conducted, adequate study that are interpreted as showing no biologically relevant evidence of chemically related deficits in reproductive parameters. A statement to the effect of "This study had no observable adverse reproductive toxicity at the highest dose tested (XXXX mg/kg/d)" should accompany the evidence statement.

Inadequate Study of Reproductive Toxicity

- Demonstrated by a study that, because of major design or performance flaws, cannot be used to determine the presence of reproductive toxicity.

Dr. Carney presented the key points to consider when applying the levels of evidence criteria and said these points would evolve. When a conclusion statement for a particular experiment is selected, consideration must be given to key factors that would extend the boundary of an individual category of evidence. Interrelationships between endpoints and the impact of the change on reproductive function must be considered. He gave as an example a decrease in pup body weight that resolves itself versus a true change in organ weight or sperm count. Also for consideration are the relative sensitivity of endpoints, normal background incidence, and specificity of effect. For certain endpoints, the statistical power is more than adequate, but for incidence, the studies are underpowered.

Factors to consider in selecting the level of evidence include: (1) increases in severity and/or prevalence as a function of dose, which generally strengthens the level of evidence; (2) the more animals affected, the stronger the evidence; however, effects on a small number of animals across multiple related endpoints should not be discounted, even in the absence of statistical significance for the individual endpoint(s); (3) malformations with low incidence should be interpreted in the context of historical controls and may be biologically important; (4) consistency of effects across generations and endpoints strengthens the level of evidence; (5) transient changes by themselves may be weaker indicators of effect than persistent changes; (6) single endpoint changes by themselves may be weaker indicators of effect than concordant effects on multiple, interrelated endpoints; (7) insights from supportive studies and reproductive findings from other *in vivo* animal studies should be drawn upon when interpreting the biological plausibility of a change; and (8) clear changes in multiple reproductive tract endpoints without functional changes are sufficient for clear evidence of reproductive toxicity. He added that the RDCWG supported using new technical approaches as they are developed.

b. BSC Discussion

Dr. Bunton said the working group meeting went remarkably well. People from different disciplines were able to take the draft criteria and case studies and come pretty close to the same conclusions. There were some discrepancies, but for the most part, applying the criteria worked.

Dr. Portier said there was a lot of discussion about statistical significance, the size of the studies NTP does, and whether they are adequately powered. The criteria do not state "statistical significance," but the working group discussed it. Five of the working group members sat on the NTP Center for the Evaluation of Risks to Human Reproduction (CERHR) Bisphenol A Panel and were familiar with reproductive toxicology studies and their statistical issues.

Dr. Mirsalis, a lead discussant, said the working groups had his utmost respect because this was a huge, almost impossible job. Carcinogenicity is pretty objective such that a layperson could look at the historical incidence table, discern a significant increase in adenomas, and determine that it meets the criteria for clear evidence. It is very cut and dried. With reproductive toxicology studies, he thought a very sophisticated group like members of the working group could look at the case studies and determine with tremendous precision the level of evidence. He did not think the public would be able understand how to apply the level of evidence criteria to the reproductive studies as easily as they could with the carcinogenicity studies. Professional judgment is necessary to use the reproductive criteria. He expressed concern that NTP would get into unresolved issues as they did with genotoxicity. He said today there are three categories for genotoxicity: plus, minus, and plus/minus. With that system of classification, vitamin C, phenobarbitol, and aflatoxin are positive. He had suggested using a scale from 0 to 100 for assessing genotoxicity outcomes. Dr. Mirsalis expressed concern that for something complicated like reproductive and developmental toxicology with many different factors, it might be hard to make a clear call. He supported linking the call to the dose level for all the criteria. He thought the criteria were as good as they were going to get. He was concerned about how the classifications would be used by regulators and the public.

Dr. Pino,[1] agreed with the RDWG's inclusion of a dose level in all the criteria. He suggested not using the terms LOAEL and NOAEL for the "some evidence" and "no evidence" categories, respectively, but just to indicate the dose of the effects. He thought developmental effects and reproductive effects discussions should be separated in the reports, e.g., the malformation discussion should be in the developmental report..

Dr. Riviere thought the doses should be included in the levels, but the term LOAEL should not be used. Dr. Crump said he did not oppose including dose levels, but that it would then make using the criteria a dose-response assessment, not a hazard assessment, according to "Red Book" standards. He added that if doses are included for one set of criteria then they should be included for all three criteria for consistency. He suggested using the term "treatment-related" rather than "dose-related," as it might be better understood.

Dr. Friedman-Jiménez thought the statement regarding LOAEL levels was quite ambiguous and uninformative. He suggested including the power, the confidence interval, or some other statistic, and the sample size to determine how much weight should be put on the evidence.

[1] Dr. Pino participated in the BSC discussion by teleconference.

Dr. Howard said the criteria are for notifying the public of the study's findings. He thought using terms like LOAEL and NOAEL, without defining them, could be misleading. He asked about including doses in the summary statements of the carcinogenesis reports. Dr. Bucher said doses are included in negative cancer studies to indicate the highest dose at which the chemical was evaluated. Dr. Howard asked why the LOAEL would be included in these criteria if they were not included in the cancer summaries. Dr. Bucher said it might be because of a perception that cancer may be linear and not have a threshold while non-cancer outcomes may be nonlinear.

Dr. Portier said the doses were included because the studies address so many different endpoints and there may be significant results in a number of them. The dose gives the reader an idea of the lowest dose NTP thinks that a clear reproductive event occurs. It is a weight of evidence concept. Dr. Howard used the example of Triclosan with a NOAEL of < 1.5 mg/kg/day, which he considered totally useless information. He said stating one dose without providing the step increases in the doses would not be informative. He thought stating both the LOAEL and NOAEL doses would be very complicated. Dr. Keith Soper said the dose levels would depend on the study design and would not be reproducible from study to study, unlike using model-based numbers. He added that NTP will need to be very clear in the use of statements like "professional judgment" and "weight of evidence across multiple factors," or they will not be comparable across studies, across time, and across compounds.

Dr. Wegman said the use of the term "insight from supportive studies" is confusing when NTP is referring to one study in using the criteria. The use of the term "dose-related" in describing "clear evidence," but not using it in "some evidence" is confusing. He asked if it means with "some evidence" that the deficits are not dose-related. He thought the statement, "Relative to clear evidence of reproductive toxicity, such effects would be characterized by greater uncertainties or weaker relationships with regard to dose, severity, magnitude, incidence, persistence and/or decreased concordance among affected endpoints," would be confusing for future cross-study comparisons. He said including the highest dose at which effects were not seen includes issues of both power and level.

Dr. Toraason said including the dose in using the criteria crosses the line into risk assessment and he thought the dose would be used inappropriately. The reader should decide how to use the dose. He also suggested wordsmithing the criteria to make them more succinct.

Dr. Vandenbergh asked if part of the intent is to improve future studies. If so, he suggested adding information, as was discussed by the CERHR Bisphenol A expert panel, such as maternal influence and power. He said just measuring fertility may cause the NTP to miss a lot of behavioral effects, in terms of reproductive mating behavior.

Dr. Toraason asked if the working group came up with the same dose levels for NOAEL and LOAEL when applying the criteria to the case studies. Dr. Carney explained that the case study examples did not include the whole report, but were just bulleted findings, so no calls were made on the doses. He thought stating the dose levels would be relatively

straightforward because the information would be in the technical report. Dr. Portier said dose levels and parameters are presented in the conclusions of the reports. The goal was to point to the dose level at which important events occur, possibly with the use of terms other than LOAEL or NOAEL. He realized that agencies will use the information however they choose, but the risk assessors would never be as familiar with the data as the panel making the decision about the dose level. Dr. Kerkvliet said including a number implies that NTP is defining the dose level in the context of the NTP study only.

Dr. Crump then agreed with Dr. Toraason that the dose should not be included because the criteria are not for hazard assessment. The dose level is a statement about one NTP study, but it will be misinterpreted as a statement about the chemical. Since there might be specific doses for each of a number of outcomes in the study, it might be difficult to identify a specific low dose relating to the study. As with cancer, there is no objective evidence that these toxicities have a threshold, so these criteria should follow the same tradition and not include a dose. Dr. Wegman agreed with Dr. Crump and said since the doses can be found in the technical report they do not need to be included in the summary statements. He thought the term "hazard" was inappropriately used in an individual study context.

Dr. Carney thought inclusion of a dose number was appropriate for an individual study because the goal is to state what happened, and at what dose level, for that study.

Dr. Soper moved that the report be accepted, with the understanding that the discussion would be captured in the minutes of the meeting. Dr. Mirsalis seconded the motion. The motion passed with 12 yes votes, 1 no vote, and 1 abstention. Dr. Friedman-Jiménez opposed the motion because he thought the dose number should be removed due to it being rife with ambiguity. He said the data would be in the technical report and should not be included in the summary statements. Dr. Wegman abstained because he considered the description in "some evidence" unable to be interpreted.

4. Levels of Evidence Criteria for NTP Developmental Toxicology Studies

a. Presentation

Dr. Carney said the same key points identified for the reproductive toxicology criteria applied to the criteria for developmental studies. The criteria are relevant to individual studies; they note the strength of evidence; there are the same caveats about negative and positive results; developmental events are intertwined in the reproductive process; and the levels are for reporting hazard only, not risk. The levels of evidence have the same five categories and report LOAEL (clear, some evidence) or NOAEL (no evidence).

Dr. Carney then explained the levels of evidence the working group drafted for evaluating developmental toxicity:

Clear Evidence of Developmental Toxicity

- Demonstrated by a dose-related[1] effect on one or more of its four elements (embryo-fetal death, structural malformations, growth retardation or functional

deficits) that is not secondary to excessive maternal toxicity. A statement to the effect of "This study has a lowest observed adverse effect level of XXXX mg/kg/d for developmental toxicity" should accompany the evidence statement.

[1]The term "dose-related" describes any dose relationship, recognizing that the treatment-related responses for some endpoints may be non-monotonic due to saturation of exposure or effect, overlapping dose-response behaviors, change in manifestation of the effect at different dose levels, or other phenomena.

Some Evidence of Developmental Toxicity

- Some evidence of developmental toxicity, relative to clear evidence, is characterized by greater uncertainties or weaker relationships with regard to dose, severity, magnitude, incidence, persistence, and/or decreased concordance among affected end points.
- A statement to the effect of "This study has a lowest observed adverse effect level of XXXX mg/kg/d for developmental toxicity" should accompany the evidence statement, except in those instances in which the "some" classification has been based on uncertainties about the dose relationship that precludes confident determination of the LOAEL.

Equivocal Evidence of Developmental Toxicity

- Demonstrated by marginal or discordant effects on developmental parameters that may or may not be related to the test article.

No Evidence of Developmental Toxicity

- Demonstrated by data from a well-conducted, adequate study that are interpreted as showing no biologically relevant evidence of chemically related effects on development. A statement to the effect of "This study had no observable adverse developmental toxicity at the highest dose tested (XXXX mg/kg/d)".

Inadequate Study of Developmental Toxicity

- Demonstrated by a study that, because of major design or performance flaws, cannot be used to determine the presence of developmental toxicity.

Dr. Carney said even though developmental toxicology has a smaller range of endpoints, it is necessary to look for interrelationships between endpoints and for the impact of the change on developmental function. Other points to consider are relative sensitivity, normal background incidence, and specificity of effect. For evaluations on the borderline between two levels, it is necessary to consider increases in severity and/or prevalence. He said thalidomide is a good example of a clear developmental toxicant, whereas ethanol is clearly a human teratogen, but in looking at just animal studies, effects occur only at very high dose levels in the presence of maternal toxicity.

A greater weight of evidence is given to (1) selective effects on the embryo/fetus/pup; (2) effects seen in many litters; (3) concordant effects; (4) statistical increases on a litter

basis; (5) large numbers of animals being affected, though effects in small numbers across multiple endpoints are important; (6) consistency of effects across generations; (7) persistent versus transient changes; (8) insights from supportive studies; and (9) similar effects observed in a second species. Other key points include that the studies should be well designed and of adequate statistical power and that new technical approaches and highly sensitive techniques need to be appropriately utilized.

b. BSC Discussion

Dr. Portier said the combination of several endpoints occurring together can move the level from "some" to "clear," so the concordant effects make it hard to discuss in terms of statistics. The combination of effects is important, not just the one event. He said it is important to mention supportive studies, e.g., maternal toxicity, and integrate it indirectly.

Dr. Mirsalis, a lead discussant, said his comment about the reproductive criteria apply to the developmental criteria. He reiterated his earlier comment supporting the inclusion of dose in the conclusion statements, including the carcinogenicity studies, because he does not consider cancer a non-threshold event. He gave the example of the delayed ossification endpoint and said he would rather have a knowledgeable NTP review panel make the call on the dose at which effects are seen than to leave it unspoken.

Dr. Pino, a lead discussant, said his comments were the same as for the reproductive criteria. He asked for clarification on the key point regarding the relationship between maternal physiology and development. Dr. Foster said frequently there is more concern when a selective effect is seen on the embryo/fetus, in the absence of effects on the dam. But one should not dismiss the effects on the embryo/fetus when there are effects on the dam. Dr. Pino said information on dose could be useful information to have and he had no objection to including it; however, he would rather not have it expressed as a NOAEL. Dr. Crump said it might be confusing if the NTP includes the dose when "some evidence," of an effect is the conclusion. It appears that NTP is reporting that they are not sure something is occurring, but it is occurring at a certain dose. He said the events are being interpreted as threshold phenomena, when there is no evidence that they are threshold or non-threshold. Dr. Carney said the language could be improved, but "some" is still a positive effect and he mentioned thalidomide. He said many of the compounds tested are not as severe, but they're still positive, and thus can have a dose level put on them. It is not to suggest a threshold, but to state what happened at what dose in the specific study. He said it is important to communicate that information.

Mr. Janzen supported including the dose number when there is a clear dose-response.

Dr. Wegman expressed confusion about the terms "statistical increase" and "absence of statistical significance" in the key points. He said the RDCWG report should specify what they mean. He thought the point regarding study design and statistical power is important and should be included in the ICWG report.

Dr. Howard said including the dose numbers could mislead readers to making erroneous conclusions. To keep the statements concise, he recommended letting the reader find out the doses at which effects occur and draw their own conclusions. Dr. Friedman-Jiménez was concerned that including dose as a single number would leave open a wide variety of interpretations, misinterpretations, and misuses. He suggested presenting more clearly defined statistics that incorporate both the size of the effect and the statistical significance or not including the dose in the summary statement.

Dr. Kerkvliet said all the comments would be taken as advisement to NTP staff. She called for a motion. Dr. Mirsalis made a motion to accept the report with the understanding that the comments were recorded and would be taken into consideration by the NTP staff in revising the criteria. Dr. Novak seconded the motion. The motion passed with 13 yes votes, 0 no votes, and 1 abstention. Dr. Wegman abstained for the same reason as for the reproductive toxicity criteria, that the description of "some evidence" is inadequate and not interpretable.

VI. DNA-based therapies

a. Presentation

Dr. Richard Irwin, NIEHS, described NTP's participation in this initiative. The FDA nominated for study biologics that contain DNA or other nucleic acids as their major components because (1) it had limited authority to require any testing other than short-term acute testing for these products despite a concern of their possible persistence, (2) the majority of the sponsors are small companies or academic institutions that lack the resources to conduct extensive in-depth evaluation of their products, (3) the FDA is unable to disseminate proprietary information on a particular vector that might be useful for another applicant, and (4) DNA-based therapies are one of the fastest growing product portfolios that the FDA has.

NTP is involved in two major activities. One is a joint study with the FDA to examine insertional mutagenesis of retroviral and lentiviral vectors into hematopoeitic stem cells. The second is collaboration with the National Institute of Dental and Cranial Research that uses the salivary gland as a target for viral transduction. Although the salivary gland is an exocrine gland, it can secrete endocrine products into the blood.

The major risk associated with the use of retrovital and lentiviral vectors is insertional mutagenesis where integration occurs randomly in the genome within transcriptionally active chromatin. No consensus sequence has been found for the site of integration of a particular vector. The long terminal repeats (LTR) of retroviral vectors contain a strong promoter enhancer that can disrupt the control of transcription of cellular genes in the vicinity of the integration site, particularly if integration occurs in the coding or control region of the gene. Ectopic expression of the gene product could interfere with host cell signaling and disrupt normal cellular processes leading to oncogene activation or tumor suppressor inactivation.

One genetic disorder, severe combined immunodeficiency syndrome (SCID), has been

the subject of a number of gene therapy clinical trials. The most common form of SCID is the X-linked form caused by mutations in the common gamma chain of cell surface receptors for several interleukins. Children born with X-linked SCID are unable to mount an immune response and without intervention die within two years. The first successful human gene therapy trial involving children with SCID was published in 2003. The bone marrow cells from 10 SCID patients were successfully transduced with a therapeutic gene, and the reinfused cells repopulated their bone marrow and expressed the therapeutic gene product that corrected the disease. Subsequently, nine of these patients presented with a lymphoproliferative disorder and leukemia and died due to insertional mutagenesis of the vector in the vicinity of the LMO2 promoter. LMO2 is a T-cell oncogene that is overexpressed in many lymphoblastic leukemias.

In 2002 a paper was published in which insertional mutagenesis was observed in a mouse gene marking trial. Gene marking is a technique used in studying cell lineages in bone marrow transplants. Bone marrow cells are exposed to a retroviral vector that expresses a cell surface protein so the cells can be monitored using flow cytometry. The protocol in this study differs from the human studies in that the transduced and expanded bone marrow cells are injected into mice that have been irradiated with 10Gy of radiation to destroy their intact bone marrow. The transduced bone marrow cells successfully reconstituted the bone marrow of the animals. These primary recipients showed no hematopoietic disorders after 10 months; however, when bone marrow from these primary recipients was injected into irradiated secondary recipients, within seven months all the mice developed leukemia that could be traced to a single primary insertion. The experiment was repeated and the same results were obtained. A subsequent study showed that not all vectors induced leukemia. This finding suggested this technique might be useful as the basis for an assay to assess insertional mutagenesis in hematopoietic stem cells and measure the effect of dose in terms of the number of vector particles administered, the multiplicity of the insertions, and the effects of changing the configuration of the transgene in the vector. If the dose could be adjusted so that there is only one insertion per cell, one could collect material to analyze for the insertion site.

The NTP is working with the FDA to validate this assay. The bone marrow donors will be C57Bl6 mice that carry the normal B allele for CD45, a cell surface protein, and the recipients will be B6SJL mice, which carry the A form of the allele. The step in which the mice are irradiated and the bone marrow destroyed is the most difficult because the animals will die if effective engraftment of the transduced cells is not obtained. Presently, the NTP is conducting two pilot studies using two different vectors: a positive control vector with intact LTRs, which is known to cause leukemia, and a second self inactivating vector where the U3 region is deleted from the LTR and inserted in the interior of the vector (SIN vector) to drive the expression of the transgene linked to green fluorescent protein. Engraftment was successful with both vectors with 50% of the bone marrow being donor bone marrow even though the SIN vector was slightly toxic. The definitive study will consist of 50 animals with two multiplicities of infections of the vectors. Tissues will be collected and DNA extracted for insertion site mapping.

The second area of investigation involves using the salivary gland as a target for viral

transduction. The objective is to conduct preclinical evaluation of the vectors to assess the toxicity associated with salivary gland transduction, persistence and duration of transgene expression, secretion of the protein into the serum, and possible systemic distribution of the vector. This approach will be useful for treating single protein deficiency diseases such as growth hormone or clotting factor deficiencies. This procedure will also be useful in patients that undergo irradiation for head and neck cancers because salivary gland function is lost. The salivary gland has a number of attractive advantages: accessibility without surgery, possibility of cannulation of the salivary gland duct without anesthesia, and limited systemic distribution of the vector as the salivary glands are encapsulated. Following cannulation and injection of the vector solution into the salivary gland, the protocol is similar to a standard toxicity study with four to five dose groups and the collection of blood and saliva on days 3, 29, 57, and 92 to measure the level of therapeutic protein. At necropsy, tissues are collected for hematology, clinical chemistry, and the presence of vector sequences.

NTP has published the results from three successful studies in which no adverse responses occurred in any of the animals and there was a dose-related increase in the secretion of the protein from the salivary gland without systemic exposure of the vector in the animals. The proteins secreted were human growth hormone, human aquaporin I from animals irradiated in the head and neck region, and a human erythropoietin transgene that produced a dose-related increase in erythropoiesis in the animals. The distribution studies in the erythropoietin study indicated that the levels of the vector were higher in males than females, which has not been observed for another vector.
NTP will be involved in the development and evaluation of models for the preclinical evaluation of vectors for other DNA containing therapies so that the information can be publicly available.

Dr. Irwin mentioned Ms. Molly Vallant, the NTP project officer for contract laboratories where these studies were conducted.

b. BSC Discussion

Dr. Vandenbergh asked whether there is a difference if the submaxillary or parotid salivary glands are cannulated, and Dr. Irwin replied that he did not know if both glands have been examined. Dr. Vandenburg said secretion of saliva is known to be androgen-dependent, particularly in the parotid gland, which is larger in males.

Dr. Mirsalis asked whether any naked DNA plasmids are being studied in the absence of a vector, and Dr. Irwin replied that naked plasmids were considered, but the NTP has not studied them. Dr. Mirsalis added that many DNA-based vaccines contain plasmids. His company has evaluated about 25 plasmid vaccines, but in most cases, the information is proprietary. Some of the naked plasmids are retained and genomic integration appears to be sequence or size-specific. It would be interesting to investigate why one of two similar-looking plasmids integrates into the genome and another does not. Dr. Irwin replied that the reason for the studies is to make a safe vector that would be clinically useful for gene therapy.

Day 1 of the meeting adjourned at 4:45PM.

November 21, 2008

Dr. Kerkvliet called the meeting to order. The BSC and public introduced themselves.

VII. Contract Concept Review on the Procurement of Mold Materials

a. Guidelines for Review

Ms. JoAnn Lewis, Office of Acquisitions at the NIEHS, briefly outlined the guidelines for the BSC regarding the discussion of research concepts. She asked the BSC to review the concept on mold materials for its overall value and for its 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 will 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.

b. Presentation

Dr. Germolec, NIEHS, outlined the concept to produce mold materials for study by the NTP. Mold was nominated and approved for study by the BSC and NTP Executive Committee about four years ago. A year ago she updated the BSC on NTP's plans to study mold due to the paucity of studies where its health effects have been investigated in a controlled environment using scientifically sound protocols.

Dr. Germolec discussed the types of mold materials and the focus for NTP's studies. The mold materials would be used to (1) study organ system toxicity following inhalation exposure, (2) develop biomarkers for exposure, and (3) assess the contributions of these organisms to health effects. Four organisms, including two *Stachybotyrus* isolates, one *Aspergillus* spp. and one *Alternaria* spp., will be studied as well as two mixtures to simulate real life scenarios. The mixtures consist of mold from a water damaged building in New Orleans and mixed cultures of mold from a damp building where human health effects have been reported.

The specific objectives of the contract are to establish a mechanism to procure the materials needed to conduct these studies. The individual organisms and the mold mixtures will be grown on relevant building materials. The NTP will supply the mold colonies to the contractor who will culture the organisms and characterize the individual cultures and mixtures to ensure the species characteristics. The contractor will confirm the viability and specific life stage of the material(s) they produce. In total, the NTP

needs 200 kg of dried material to conduct the inhalation study. The NTP working with the NIEHS Acquisitions Branch has determined that there are interested parties who will respond to a request for contract. The statement of work is completed and if the BSC approves the concept, a contractor will be sought.

c. BSC Discussion

Mr. Janzen, BSC lead discussant, provided comments on the concept review document. The ability to procure molds that are consistent and reproducible is important to the needs of this program. NIEHS does not have the requisite facilities to produce the material, thus, outsourcing the production of the materials is the best option. The technology is available to grow and culture mold consistently in large volumes. The issue will be proper quality control of the growth conditions and establishing a quality plan with the producer that will ensure reproducible materials over a long period of time. The vendor must have a quality control plan in place, provide details on the conditions for the growth of the organisms, and establish a mechanism to transfer the information on the methodology. It is not uncommon for fungal colonies to show phenotypic drift and alterations in metabolite patterns, but the NTP plans to have these characteristics monitored to ensure that the molds produced mimic the wild type's growth patterns. He proposed approval of the concept so the NTP could move forward.

Dr. Portier said quality assessment (QA) and quality control (QC) are challenges and asked whether the mixed colony would be of the same composition each time it is grown. Dr. Germolec replied that the mixture would not be maintained as a live culture over a long period of time thus, there would be less drift in the organisms. The material would be grown in bulk and lyophilized. The toxicology study will include exposure to dry dust. The NTP will characterize the mixture at the beginning, end, and during the study to identify the organisms to which the animals are exposed. The statement of work will include requirements for good QA and QC, and the NTP will monitor the materials for their consistency throughout the study including spore counts, fragments in air, and toxicological products produced by the mold that may grow in the bedding.

Dr. Howard said not many companies or laboratories would be able to generate such large quantities of mold. He wondered whether the NTP would provide samples to extramural laboratories so they could compare these organisms with their cultures. Dr. Germolec said in 2006 the NTP sought input from a group of scientists familiar with the challenges in mold toxicology. They thought the NTP would be able to obtain *Stachybotrus* cultures from academic laboratories for comparison. The NTP would be willing to be a resource and to compare its results with academic laboratory samples and to share their methodologies to characterize samples.

Dr. Mirsalis asked whether the mold would be produced under good laboratory practices (GLP) or good manufacturing processes (GMP) equivalent to those used in drug production and whether NTP had undertaken preliminary work to ensure that the mold can be aerosolized. Dr. Germolec said she hoped GLP standards would be used, but it would depend on the response from the offerers. The first part of the contract would be a

feasibility study prior to starting bulk production to ensure that the mold can be generated, aerosolized, and characterized.

Mr. Janzen asked about the chemical characterization of the fungi, if the wild-type culture of the mold would be characterized, and how the NTP would respond if its evaluation of the culture differed from the vendor. Dr. Germolec responded that the NTP would characterize the mold used in the exposures before, during, and after the cultures were administered. Different cultures of *Stachybotyrus* have different metabolite profiles. Mr. Janzen then asked if the mold collected from the building with "sick building syndrome" was characterized immediately after collection and again after growth in culture. Dr. Germolec said the selected mixtures would be well characterized as to the organisms in the culture and their potential health effects.

Mr. Janzen entertained a motion to approve the concept, Dr. Novak seconded it, and it was accepted unanimously with 12 yes votes.

VIII. Toxicogenomics Study

a. Presentation

Dr. Irwin provided an overview of NTP's studies on toxicogenomics. The basic responsibility of the toxicogenomics faculty is to develop and execute a strategy for the implementation of toxicogenomics; develop, review, and approve research proposals; and serve as a resource for NTP scientists.

He described the three areas of investigation by NTP. The first area is to examine the basic issues related to the design of toxicogenomics studies. The extent of normal variation among animals used in NTP studies is being monitored because studies are conducted in many laboratories. Understanding the effect of various study parameters, including animal husbandry, timing of tissue collection, Circadian rhythm, periodicity of female cycles, age-related effects, and variation between laboratories is important. The liver is the only organ that has been studied. The second area is incorporation of toxicogenomic studies into pre-chronic studies to help elucidate mechanisms of toxicity and refine the methods for dose setting for chronic studies. The third area is identification of gene expression patterns associated with cancer, specifically hepatocarcinogenicity.

Dr. Irwin described completed studies on the sources of variability in toxicogenomic studies.

Extent of variation in animals

Six rats per group were given 0.5% methylcellulose by gavage. Samples were collected after 6, 18, 24, or 48 hours and RNA was isolated from the liver for transcript profiling. There was detectable animal-to-animal variation in the transcript profiles, but no single transcript was significantly expressed in the rats. The basal gene expression in the male rats did not exhibit statistically significant differences from one another, thus, necropsy is well controlled, uniform, and reproducible within a study.

Effect of circadian rhythm

Circadian rhythm can impact the hepatic transcriptome. This study was designed to determine which transcripts were affected by circadian rhythm over a three-day period in two groups of animals. One group was maintained on a normal daylight cycle and a second on a light reversal cycle. Animals received a dose of methylcellulose and six animals were necropsied after 6, 18, 24, or 48 hours. RNA from each animal on the normal daylight regimen was hybridized with a pool of RNA from the animals on the reversed light cycle so that any change in the circadian rhythm would be maximized. Pooled RNA from each of the four time points was hybridized to the unit reference sample so that each time point would be compared to the same standard.

The study showed that 972 transcripts expressed in the 12-hour light reversal cycle group included circadian genes, clock genes, clock control genes, and a number of genes related to intermediary metabolism. Using a Fourier analysis, 1300 transcripts were found that exhibited a cyclic pattern of expression and 200 were in common with the 12-hour group. These findings indicated that the circadian rhythm has a pronounced effect on the hepatic transcriptome, thus groups of animals that will be compared need to be necropsied within a 3-4 hour time window to avoid confounding the transcript profile with the effects of circadian rhythm.

Effect of age

Age-related changes of the hepatic transcriptome were studied in sentinel animals from NTP chronic studies after 6, 12, or 18 months on a control diet; the animals began the diet at 6 weeks of age. The animals were necropsied at 32, 58, or 84 weeks of age, respectively. Isolated hepatic RNA from all animals for each age group was hybridized and pooled.

A heat map showed hierarchal clustering (which is the partitioning of data into a cluster such that each cluster is more closely related to each other than data assigned to different clusters; hierarchal clustering is the ordering of clusters using a top down or bottom up approach) within the 6-month-old animals and 18-month-old animals while the 12 month-old animals were represented in both groups. A significant age-related effect was noted in these transcripts, thus it is important to use age-matched and time-matched controls.

Future Studies

Effect of estrous cycle

The NTP is studying the impact of the estrous cycle on the hepatic transcriptome in female animals using an approach similar to that used for the circadian rhythm study. This study will determine the relative impact of different stages of the estrous cycle on the hepatic transcriptome and provide a basis for determining the appropriate stage and time window during which treated and control animals must be necropsied. This study should aid the interpretation of sex differences in the liver in response to chemical exposures.

Study of the microcystin class of algal toxins
About 60 different microcystins have been identified in surface waters using a phosphatase inhibition assay. Microcystins are difficult to obtain in large amounts and impossible to study using conventional toxicology methods. The NTP will study whether different microcystins and microcystin mixtures exhibit similar hepatic toxicity (based on pathology, clinical chemistry, toxicogenomics) at comparable toxic equivalent doses. ,

Other ongoing studies are pre-chronic studies of cardiotoxins where heart RNA will be collected for transcriptome profiling. It is hoped that transcriptome profiling will aid in the interpretation of toxic responses and improve hazard identification.

b. BSC Discussion
Dr. Mirsalis was impressed with the studies and pointed out their importance. He was pleased the NTP is evaluating the estrous cycles and circadian rhythms on the hepatic transcriptome.

Dr. Toraason said the relevance of these temporal parameters and estrous cycle effects are important, but their impacts on sampling were known before the studies were implemented. Dr. Irwin agreed, but said the significance of the impacts was unknown. These data will aid in the design of studies and in eliminating these parameters as confounders. Dr. Toraason said the temporal studies provide an opportunity to evaluate susceptibility changes and gender effects over time, and to make predictions about toxic responses and develop hypotheses to test.

Dr. Carney said toxicogenomic profiling would seem to be useful across a class of chemicals, such as the glycol ethers. If there are good data on a prototypic chemical in a class, this information could be used to determine if a candidate chemical falls within the class and might avoid additional testing.

Dr. Kerkvliet asked about the controls used in the age study for the pooled RNA. Dr. Irwin said there were 63 animals of different ages: 6, 12, and 18 months. RNA samples were taken from each individual animal, pooled, and used as a control against which all the samples were compared.

Dr. Novak asked if a reference standard was used so cross comparisons could be used for future studies, and whether the analysis corrected for dye bias. Dr. Irwin replied that the analysis corrected for dye bias and used this approach to develop a control because they had not determined how useful a reference standard was.

Dr. Friedman-Jiménez asked how the time points of 4, 10, 16, and 22 hours were selected for the circadian rhythm study. Dr. Irwin replied that those times corresponded to 6, 12, 18, and 24 hours, respectively, after the initial dose. Dr. Irwin replied that these times were selected to maximize differences and because these time points are used in other NTP studies.

Dr. Kerkvliet asked about the difficulty of superimposing the estrous cycle time points

with the circadian rhythm when sampling for a biomarker of effect, if sampling occurs in the middle of the night. Dr. Irwin replied that selection of a specific time point in the estrous cycle must be justified and the time point chosen must be matched between the control and treated animals. Dr. Walker said the design team considers both the circadian rhythm and estrous cycle when designing studies in females.

IX. Predicting Hepatocarcinogenic Potential of Alkoxypropenyl Benzene Derivatives using Toxicogenomics

a. Presentation

Dr. Scott Auerbach, a NIEHS post-doctoral trainee, made the presentation. His study investigated the hypothesis that hepatic pattern-recognition models trained on hepatic gene expression profiles induced by hepatocarcinogens and non-carcinogens can identify the alkoxypropenyl benzene derivatives that pose a significant hepatocarcinogenic hazard.

He defined a number of terms. (1) Supervised machine learning is the computational method used to generate pattern recognition models. One employs prior knowledge about the chemical in order to search for genes that correlate with a disease state or the response to a carcinogen or noncarcinogen. (2) Training data are mRNA expression data used to train the pattern recognition models. (3) The test data are the mRNA expression data not used for training the models but used to independently evaluate the performance of the models. (4) Cross-validation is the use of the training data to create the model and classify the samples that were used to train the model. (5) Independent validation is the classification of samples not used to train the model. (6) An optimal model is a pattern-recognition model that achieves 0% (or as close to 0% as possible) cross-validation error with a minimum number of genes. An optimal model is identified to allow for easier biological interpretation of the model and increase its generalizability, so that there is a greater chance of properly identifying another chemical as a carcinogen or non-carcinogen in later studies.

The alkoxypropenyl benzene derivatives are a large class of naturally occurring or synthetic compounds used in fragrances and flavoring agents; a number are approved as additives to human foods. A limited number have been studied in 2-year carcinogenicity bioassays; some including methyleugenol, estragole, and safrole cause hepatic cancer in male rats.

Prioritization for testing is necessary for a large class such as the alkoxypropenyl benzene derivatives. To create training data, male F344 rats were dosed for 2, 14, or 90 days with known carcinogens or known non-carcinogens. The doses chosen for the carcinogens were based on an expectation that they would elicit tumors in 40% of the animals by 2 years. The doses of the non-carcinogens were chosen based on an assumption that they would elicit gene expression. Hepatic mRNA was isolated and the expression of 41,000 genes measured in each sample. The mRNA expression levels were used in a supervised machine learning method to create and optimize the carcinogenicity prediction models based on a single or combination of exposure durations. The model was evaluated using

M-fold cross-validation. In this method, meaningful information from nine of 10 treatments was extracted and then a model was created for the deleted treatment. This is done 10 times, once for each of the 10 treatments. The models generated using the training set were then used to classify gene expression from males F344 rats gavage-dosed with alkoxy propenyl benzenes for 2, 14, or 90 days at dose levels of 0.2 or 2 mmols/kg/day.

He discussed the seven characteristics of the optimal pattern-recognition models that were identified based on the training data from the 2-day, 14-day, 90-day, 2+14-day, 2+90-day, 14+90-day, and 2+14+90-day studies. All optimal models with the exception of the 2+14-day model achieved 0% error by cross-validation with the number of genes per optimal model ranging from 3 to 59. All the models classified the 90-day test data with higher accuracy than either the 2 or 14-day test data. All models performed with similar accuracy when predicting the 90-day test data; therefore, an assembly model approach was used to classify gene expression produced by the individual alkoxypropenyl benzenes.

There are seven models with 10 animals /dose group resulting in 70 predictions/dose group. In the case of isoeugenol, all of the models classified it as a non-carcinogen. The model identified methyleugenol, at both the low and high dose, as being carcinogenic. At the low dose, 45% to 65% of the samples suggested that estragole is carcinogenic, but at the high dose every sample classified it as a carcinogen. Similar results were obtained with safrole. Anethole at low and high doses did not produce gene expression patterns indicative of carcinogenic activity. There was a notable carcinogenic signature with myristicin at the high dose. Isosafrole showed some indication of carcinogenicity at the low dose, but a notable carcinogenic profile at the high dose. In a bioassay of isosafrole with doses of 10,000 ppm, 5,000 ppm, and 2,500 ppm, all the animals at the highest dose died, while there was an uptrend of hepatocellular carcinoma at 5,000 ppm. The dose used in the toxicogenomic study was equivalent to 6,000 ppm. This study identified a carcinogenic profile at a dose where tumors might be expected.

He listed the genes informative to these models. A Venn diagram showed the 3, 6, and 15 genes informative to the 2, 14, and 90 day-models, respectively. Many of the genes populate pathways related to hepatocarcinogenesis, Ah receptor activation, DNA damage response, tissue regeneration, cell migration, and the cell cycle. The accuracy of the prediction for test data improved with increasing exposure duration probably because short durations of exposure for 2 or 14 days to weak carcinogen/dose combinations failed to elicit gene expression changes reflective of carcinogenic activity. Dr. Auerbach used expression of the Wwox gene to illustrate the effect of exposure duration. Its expression is decreased rapidly following treatment with high doses of carcinogens, but this response is only seen after 90 days of exposure if the substance is a weak carcinogen.

Myristicin and isosafrole should be given higher priority relative to other members of this class for testing in the carcinogenicity bioassay. He predicted that isosafrole and myristicin, tested at a 2 mmoles/kg/day day by corn oil gavage in F344 male rats, should result in a significant increase in hepatic cancer.

Highly accurate hepatocarcinogenicity prediction models can be generated from hepatic gene expression changes gleaned from rats exposed for as little as 2 days to highly carcinogenic chemical/dose combinations. Models built on 2-day exposure data are equally as accurate as models based on 90-day data; however, weakly carcinogenic chemical/dose combinations require longer exposure durations to manifest genomic changes indicative of carcinogenic activity. Based on the above findings, Dr. Auerbach concluded that data from longer exposure durations (90 days or greater) should be evaluated when using gene expression-based data to classify chemicals with unknown carcinogenic potency in order to avoid false negative predictions.

The chemicals used in the training data act by a limited number of mechanisms increasing the chance that agents acting by different mechanisms might be misclassified as non-carcinogens. It will be important to study more chemicals with varied mechanisms of action and to target other tissues in future studies. The 90-day window of exposure may be enough time to produce gene expression changes that are more universally related to carcinogenesis.

The present models do not address potency or dose-response and are limited to male F344/N rat liver. The models need to be validated across sexes, strains, and species. More models need to be created using gene expression from other common target organ systems.

b. BSC Discussion

Dr. Mirsalis asked Dr. Auerbach how he knew he did not create a predictive model for hepatotoxicity rather than for carcinogenicity, since many of the compounds tested are hepatoxins as well as hepatocarcinogens. The training set appears to be accurate for predicting genotoxic rat liver carcinogens. He wondered how this approach would fare with non-genotoxic carcinogens such as phthalate esters or chlorinated solvents that operate via a different mechanism. Dr. Auerbach replied that 2,4-dibromoanthraquinone does not appear to be genotoxic, but it upregulates the same genes that are upregulated by DNA-damaging agents related to carcinogenesis. For example, Adam8 is a gene involved in the invasiveness of cancer; 2,4-dibromoanthraquinone and other DNA-damaging agents induce it. In addition, by including acetaminophen as one of the non-carcinogens at doses that are hepatotoxic alleviates the concern that the models are simply identifying hepatotoxins as carcinogens.

Dr. Novak asked whether the tissue was stained for genes indicative of progression and whether primary hepatocytes could predict similar results to those obtained *in vivo***.** Dr. Auerbach replied that specific staining was not done, but no compounds except aflatoxin B1 caused notable changes in pathology or clinical chemistry. He added that the usefulness of *in vitro* data rather than *in vivo* data was debated at a recent HESI conference. *In vitro* systems are useful to determine the mechanism of action and classification of DNA-damaging agents, but not to create models to identify pathological endpoints such as cancer.

Dr. Miraslis said it would be difficult to develop a model with the alkoxypropenyl

benzenes from *in vitro* data. Safrole is a potent carcinogen that is consistently negative in genotoxic assays. He suggested that 28 days, as opposed to 90 days might be long enough for the study duration. He recalled that it takes a couple of weeks of dosing before safrole induces its own metabolism and DNA adducts are found, which could make the studies of shorter duration and less expensive.

Dr. Howard congratulated Dr. Auerbach and asked whether he knew if circadian rhythm impacted the genes involved in these models. If the genes linked to carcinogenic outcomes are not affected by circadian rhythms then archived tissues from previous studies could be analyzed to broaden the database. Dr. Auerbach replied that he did not know if circadian rhythm affects these genes.

Dr. Cattley was curious about the identification of genes as universal markers along the pathway to carcinogenesis; e.g., GSTπ is used as a pre-cancerous marker in the liver. The question is whether RNA changes lead to protein level changes and if they do, whether there are subpopulations of hepatic cells over expressing Adam8 at an early stage in cancer progression. This would support a hypothesis of the universality of this marker for carcinogenesis for different agents.

Dr. Soper said the question is not how well the classifiers work alone, but what improvement they provide over and above other predictive assays such as Ames. One might improve the validation of these classifiers, if the biology of the system were used to trim the set down rather than taking the features picked by the machine learning. Frequently, many genes are highly correlated and one can select the ones that are biologically relevant. Cross-classification is much better than looking at classification error rates on the training sets. He suggested testing the model with a larger sample of compounds for which the hepatocarcinogenicity is known. Dr. Auerbach said the NTP is planning to extend this study to another 20 chemicals. He added that the NTP has a large number of 90-day samples at the NTP Archives, and he hopes he can extract intact RNA from the formalin fixed tissues.

Dr. Toraason congratulated Dr. Auerbach on his progress so far and asked whether this approach would be incorporated into the bioassay program.

Dr. Nigel Walker asked for input from the BSC regarding what NTP's strategy should be for the next five years. Presently, the NTP is studying variability models, predictive models, class studies, and individual chemicals and could invest many dollars in predictive models for the kidney and lung and other tissues. He asked whether variability studies would be of more value than the class studies. Another issue is how to deal with the data that result from these studies. He asked the BSC to consider how it might react to an NTP Technical Report based on Dr. Auerbach's data predicting that isosafrole is reasonably anticipated as being a male rodent carcinogen. As the program moves further away from evaluating effects in rat liver at 90 days to *in vitro* studies, the interpretation will be more difficult. This is a test case for the use of predictions.

Dr. Riviere said he thought it would be useful to evaluate a number of different tissues,

but for now the NTP should concentrate on the liver. The NTP has a very small data set for this type of pattern recognition that should be expanded to make the system more robust. Second, it is important to test the hypothesis that there is some common gene that is up-or down-regulated across the different mechanisms of carcinogenesis. He agreed with Dr. Mirsalis about differentiating between hepatotoxins and carcinogens.

Dr. Novak asked if the NTP planned to study micro-RNAs, and Dr. Walker responded that this might be considered in the future. Dr. Auerbach said micro-RNA array technology is in its infancy with respect to interpretability. Some studies have shown promise for classifying tumor stages particularly in highly dedifferentiated tumors. Dr. Novak added that either micro-RNA profiling or real-time PCR could be done on total RNA samples to evaluate selected targets implicated in some cancers. The hepatic RNA samples treated with this class of compounds could be used as a test case.

Dr. Robins congratulated Dr. Auerbach on his work and future plans. She said switching to mice would be more cost-effective because they are smaller, cost less to feed, and develop some conditions more rapidly than rats. The advantage of mice is the better-characterized genetic variability and engineered models. She suggested that humanized mouse models would be closer to bridging to human disease. Dr. Auerbach replied that this is an excellent suggestion, but rats were chosen because mouse hepatic tumors are problematic since 40% of most control mice strains succumb to liver tumors. Dr. Robins added that in an aging study in a specific pathogen free (SPF) facility, animals die due to cancer. Males usually die of lung or liver cancer whereas females succumb to a broader spectrum of cancers hence support for using female mice. Dr. Irwin said Dr. French leads an initiative examining genetic susceptibility in different strains of mice and presently the NTP is designing an aging study in mice.

Dr. Mirsalis advised against studying mouse liver as mouse liver tumors have no relevance to human health while rat liver tumors do. In response to Dr. Walker's earlier question about predicting carcinogenicity based on genomic studies, he said if bioassays are planned with anethole, isosafrole, and myristicin and the predictions are right, that information would be very powerful. With these data, the BSC might accept genomic data in the future because of cost, time, and energy savings. He wondered how the FDA and EPA would react to receiving such a dataset.

Dr. Howard said genomics is a very active area in the FDA and industry is submitting genomic data. He said a large percentage of drugs that have potential efficacy fail because of liver toxicity. The utility of this predictive model would have a tremendous impact on future drug development if it lowers the expense of testing a compound. He thought the FDA would be unwilling to accept a predictive model as evidence that a drug is ready for a human trial. He said expanding on studies in the liver has great merit and would provide more information and ultimately regulatory acceptance of this method as a predictive tool.

Dr. Irwin added that the NTP views the liver as a sentinel organ even if it is not the primary target tissue. All absorbed compounds pass through the liver and if a compound

is toxic, there would be a response that could be detected via genomics.

Dr. Novak said if one could relate pathology from biopsies to genomic data it might have ramifications in terms of human diseases. Studying the liver is becoming increasingly important because of the incidence of chronic nonalcoholic liver disease progressing to nonalcoholic steatohepatitis and then to hepatocelluar carcinoma. This scenario is increasing because of the increase in metabolic syndrome that exacerbates the situation.

Dr. Crump said the NTP should consider how this genomic information could replace the bioassay and not just for setting priorities. He thought the BSC should look favorably on the conclusion of a Technical Report based on genomic data because the two-year bioassay is overrated as a predictor of human carcinogenicity. A few years ago he and Dr. Dan Krewski did a meta analysis of the bioassay and predicted there are probably many weak carcinogens in the database that were not declared to be positive. Given the problems of site specificity, extrapolating from animals to humans, and dose response, the bioassay is a signal and is limited in the information it provides regarding human risks. If a similar signal can be obtained from genomics with fewer animals and at less expense, a greater number of chemicals could be studied. The NTP should consider these signals in terms of setting standards for human risks and not consider the two-year bioassay as the gold standard.

Dr. Friedman-Jiménez asked whether the program is considering testing the 40 known human carcinogens in these studies because it would be more relevant to human health than comparing to an intermediate standard such as animal carcinogenicity. This data would be informative as to how well these methods perform against known human carcinogens. Dr. Irwin replied that NTP has included some of these carcinogens in the next group of chemicals the program plans to study.

Dr. Howard liked Dr. Mirsalis's suggestion of splitting the studies into two tracks: one to predict toxicity and the second to predict carcinogenicity. If the genomic findings from the two outcomes overlap that is fine, but if the predictors differ that would be better. He emphasized that the NTP must consider toxicity as one of the endpoints.

Dr. Toraason thought the NTP should invest in this program in financial and human resources. This approach would be readily accepted by NIOSH especially if dosimetry were addressed. He was intrigued with the differential response in the 14-day versus 90-day studies. This technology is more likely to be predictive than a single endpoint in a high throughput screening assay.

Dr. Cattley said if the NTP has limited resources, it should control for variability, which is predictable. Understanding carcinogenesis is more important than understanding variability, and it is not necessary to determine the contribution of each parameter.

Dr. Portier disagreed with Dr. Cattley. From a statistical point of view, it is important to understand any uncertainties and confounding factors in the experimental design. He thought more studies on laying down a framework to identify the genes that are known to

be important for carcinogenicity as well as those connected with the diurnal cycle would be necessary to proceed with the next step. Until the confounders are determined the data would be questioned.

Dr. Walker added that the NTP changed the rat strain recently and dropped the inbred Fisher 344 rat for an outbred strain. One of the concerns with any strain is variability and drift over time. One use of genomic technology might be to monitor the strains over time.

Dr. Bucher said the NTP is using three different approaches to assess human predictivity. The first is the standard toxicological model where chemicals produce pathological changes in a particular animal knowing that the species and strain influence the outcome. The second is the use of animal genetics to understand the pathology that ensues in a particular strain using the host susceptibility approach. The third is studying genetic changes to identify critical pathways that are involved in either neoplastic or non-neoplastic pathologies. An alternative is to abandon the pathological changes in the animals and go directly to the high throughput screening (HTS) activities to identify those critical targets in very short-term assays. Presently, the NTP is tackling all these approaches simultaneously, which makes it very difficult to decide where to allocate the limited dollars and brainpower. The NTP roadmap, "Toxicology in the 21st Century," touches on all these areas. He appreciated the advice of the BSC as to how the program should proceed.

X. High Throughput Screening

a. Presentation

Dr. Raymond Tice, NIEHS, reminded the BSC that the High Throughput Screening (HTS) Initiative is part of the NTP Roadmap, "Toxicology in the 21st Century: The Role of the National Toxicology Program" (http://ntp.niehs.nih.gov/go/vision). The NTP's vision is "to support the evolution of toxicology from a predominantly observational science at the level of disease-specific models to a predominantly predictive science focused upon a broad array of target specific, mechanism-based, biological observations". In late 2007, as part of its realignment, the NTP formed the Biomolecular Screening Branch to direct its HTS Initiative; this branch also includes the *Caenorhabditis elegans* screening core.

The overarching goals of the HTS Initiative are to (1) prioritize chemicals for further in-depth toxicological evaluation, (2) identify mechanisms of toxicity, and (3) develop predictive models for *in vivo* biological response in humans. The short-term goals are to (1) develop tools and approaches to characterize and probe toxicity pathways, (2) develop capabilities for including hepatic metabolism in *in vitro* cell-based assays, (3) develop and populate relational databases, and (4) prioritize chemicals for more in-depth targeted testing. The long-term goals are to (1) incorporate *in vitro* 3-D organ/tissue models into the bimolecular screening strategy, (2) inform cross-species and low-dose extrapolation, (3) continue to evaluate the genetic basis for variability in sensitivity to toxicants, (4) develop predictive models for human diseases, and (5) establish a role for bimolecular

screening in regulatory science via the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). ICCVAM recently published a five-year plan, with HTS as a component.

Dr. Tice then mentioned the relationship between the HTS Initiative and a recent National Academy of Science report, "Toxicity Testing in the 21st Century: A Vision and a Strategy," that envisions a not too distant future in which virtually all routine toxicity testing would be conducted *in vitro* using human cells by evaluating perturbations of cellular responses in a suite of toxicity pathways with high throughput robotic assisted methodologies.

In support of this initiative, the National Human Genome Research Institute (NHGRI), NIEHS, and EPA signed the memorandum of understanding (MoU), "High-Throughput Screening, Toxicity Pathway Profiling and Biological Interpretation of Findings," in February 2008. As a result of the MOU, participants from the NTP, the NIH Chemical Genomics Center (NCGC), and EPA's National Center for Computational Toxicology formed the "Tox 21 community" (for toxicology testing in the 21st century). There are four focus groups within the Tox 21 community. The Pathway Assays Group works to identify key toxicity pathways, prioritize assays for use at the NCGC, identify assay gaps, develop methods for incorporating hepatic metabolism into *in vitro* assays, and evaluate human and rodent genetic variability in response to toxicants and cell-to-cell interactions. The Compounds Group will identify a library of ~10,000 compounds with known structures for testing at the NCGC. This group will also establish procedures for determining the identity, purity, and stability of each compound and will establish a library of water-soluble compounds and a library of mixtures. The Bioinformatics Group evaluates patterns of response and the relationship of these responses to adverse health outcomes in experimental animals and humans. They also evaluate the consistency of responses within assays and across related endpoints. The fourth group, the Targeted Testing Group, will focus on prioritizing substances for more complex testing in alternative assay platforms or species (e.g., *C. elegans*, zebrafish).

The three partners in the Tox 21 community are committed to making all the data publicly accessible. This will be accomplished by placing all data generated at the NCGC in NIEHS/NTP's database known as CEBS (Chemical Effects in Biological Systems), EPA's ACToR, and PubChem.

Dr. Tice explained the assay requirements for the NCGC. As a first step, the NTP provided the NCGC with a library of 1408 compounds, consisting of 1353 unique compounds and 55 in duplicate. The EPA also provided an initial library of 1408 compounds, with about a 400-compound overlap between the two libraries. Most of NTP's library consisted of industrial compounds, with some pesticides and natural products, and focused on compounds tested in various NTP assays.

Dr. Tice mentioned the first NTP manuscript on HTS (Xia, M., Huang, R., Witt, K.L., Southall, N., Fostel, J., Cho, M-H., Jadhav, A., Smith, C.S., Inglese, J., Portier, C.J., Tice,

R.R., Austin, C.P. (2008). Compound Cytotoxicity Profiling Using Quantitative High-Throughput Screening. *Environ Hlth Perspect* **116** 284-291) which describes a cytotoxicity assay known as the CellTiterGlo™ in which 13 different cell types (9 human and 4 rodent in origin) were used to test the NTP library for cytotoxicity. Some of these cells were primary cells and others transformed cell lines. Other assays used at the NCGC are a lactic dehydrogenase release assay, proteolysis release assay, p53 assay, assays for caspase 3/7, 8, and 9, and an assay for DNA damage based on differential cytotoxicity in chicken lymphoblastoid cell lines deficient in different DNA repair pathways. The NTP library has also been tested in assays measuring agonist and antagonist activity in 10 different nuclear receptors.

Dr. Tice highlighted the potential contributions of EPA's ToxCast™ program to the MoU. ToxCast™ was formed to address chemical screening and prioritization at EPA and to evaluate the comprehensive use of HTS to generate biological and predictive fingerprints. ToxCast™ is a phased program; with the first phase testing 320 mostly pesticide active compounds in ~550 different *in vitro* and lower organism assays. A data analysis summit to discuss the results of this phase is scheduled for May 14-15, 2009, at the EPA. The 320 compounds in Toxcast™ Phase I have been tested in *C. elegans* in a growth assay.

The Tox21 partners are identifying key toxicity pathways using toxicogenomic data, human disease genetic associations, and information provided by contract organizations. Dr. Tice described the vendor meeting held September 11- 12, 2008, where companies were invited to provide information on critical toxicity pathways and useful molecular targets, as well as technologies and assay systems that might be used for HTS. Representatives from 27 organizations (mostly commercial testing companies) attended.

Mr. Janzen attended the meeting and presented brief comments to the BSC. He said the meeting was extremely well organized and the number of companies attending was impressive. Many of the companies have developed techniques to prioritize samples from drug discovery prior to them entering the regulated environment. The companies were interested in finding out how useful their techniques could be in the area of toxicology. One outcome of the meeting is that it raised the profile of toxicology in the HTS arena.

Dr. Tice briefly discussed the Tox 21 compound library under development. The EPA has evaluated the "universe of compounds" from available databases. They found 8,000 compounds with structures and 7,000 with plausible physical chemistry information. The NCGC has compiled a library of about 3,000 approved drugs that are either available commercially or can be synthesized. These compounds are being placed into a common library for testing at the NCGC and subsets of these compounds will be tested in Phase II of ToxCast™.

Dr. Tice mentioned related activities including an NIEHS Small Business Innovation Research Program (SBIR)/Small Business Technology Transfer Program (STR) on predictive test systems for safety evaluation that includes HTS and systems biology, the EU 7th Framework Program calling for proposals in high throughput research and

systems biology areas, and the EU program "Screening Methods for Assessing the Toxicological and Eco-toxicological Properties of Chemicals."

Dr. Tice closed by noting that NTP's expectations for the 21st century are to continue to refine traditional methods and to develop new methods that provide basic toxicology information for public health. There should be an effort to reconcile results from the new data-rich techniques with the existing testing information for conceptual validation, and then to develop approaches to accomplish formal validation of the new methods for human hazard and risk estimation.

b. BSC Discussion

Dr. Crump asked about the level of effort for the HTS initiative. Dr. Tice said the Biomolecular Screening Branch is relatively new and is expected to grow in terms of staff over the next year. Meanwhile, the effort at the NCGC is being funded through an Interagency Agreement and is well staffed.

Dr. Kerkvliet asked about the broad scope of the initiative. Dr. Tice replied that besides HTS data, toxicogenomic data, and other information such as Ames test data must be considered. High throughput testing has limitations because it does not consider interactions between cells in a tissue or multiple chemical interactions that may affect different pathways and exacerbate an adverse effect.

Both Drs. Soper and Crump thought the work exciting with great potential. Dr. Soper recommended NTP hire well qualified personnel that are committed to these areas and who can devote their time to carefully deciphering the data.

Dr. Crump asked how the NTP proposed to evaluate dose-response in these assays as he assumed the objective is to predict an outcome at a dose relevant to humans. He enquired about the variability within repeat experiments using the same doses. Dr. Tice replies that the maximum concentration used in the HTS assays is ~100 µM. Reproducibility is evaluated by considering the replication of duplicate compounds within and across plates.

Dr. John Bucher said the NTP is now in an exploratory mode and trying a number of different approaches to assess which ones will be the most informative. There are numerous directions the NTP can take, and an enormous number of variables need to be considered. The program has to determine which of these directions will provide the needed data. He thanked the BSC for their thoughtful comments on these new initiatives noting that the NTP relies on their advice to guide its future.

XI. Adjournment

The meeting adjourned at 2:30 PM.