The Evolution of the Science of Immunotoxicology: From a Mechanistic Foundation to Regulatory Applications

Dr. M. Holsapple

(LSI Health and Environmental Sciences Institute, USA)

THE EVOLUTION OF THE SCIENCE OF IMMUNOTOXICOLOGY: FROM A MECHANISTIC FOUNDATION TO REGULATORY APPLICATIONS.

Michael P. Holsapple, Ph.D.
Executive Director
ILSI Health and Environmental Sciences Institute, USA
Tel: 202-659-3306
Fax: 202-659-3617
Email: mholsapple@ilsi.org

The overall objective of this lecture will be to present an overview of the evolution of the science of immunotoxicology by highlighting a few key milestones, namely a solid mechanistic foundation and recent applications in a regulatory context. Neither time nor space will allow a discussion of the full extent of the evolution of immunotoxicology from a historical perspective.

The term, "**evolution**", is defined in the dictionary as "*A gradual process in which something changes into a different and usually more complex or better form.*" Suffice it to say that the growth of immunotoxicology has certainly been a "*gradual process*" over the last 25+ years due in large part to the efforts of key scientists who have insisted that this effort be based on both a solid foundation and a logical progression. There is also no doubt that immunotoxicology today is definitely "*different*" and "*more complex*" than it was even a few years ago. The last point from the definition of "**evolution**", that immunotoxicology is in a "*better form*" today than it was a few years ago, could probably be argued. There would be those who would argue that there are still no 'real world' examples of immunotoxicology. There would be others who would insist that there is no reason to exclude the potential that the immune system is a legitimate target organ. Ultimately, one proof that immunotoxicology has matured as a science lies in the attention being given it by the regulatory community, and the recognition by both the regulators and the regulated industries that potential effects of chemicals on the immune system should be evaluated.

As such, it is very appropriate to include a lecture on immunotoxicology at a symposium entitled, "**Toxicity studies in development of pharmaceutical and agrochemical products: Practical issues in GLP compliance.**" As the focus of this

lecture will be to present an update on immunotoxicity test guidelines from the U.S. Environmental Protection Agency (EPA), the information will pertain to the development of agrochemicals in the U.S. It is important to emphasize that the Food and Drug Administration (FDA) in the U.S. has its own efforts underway to update their approaches to the determination of the immunotoxic potential of drugs. It is also important to emphasize that this lecture will not address issues associated with how Good Laboratory Practices (GLP) can be integrated into the proposed immunotoxicity test guidelines. The rationale for the last point is that it was deemed to be more important to illustrate that the immunotoxicity test guidelines are mostly based on a solid 'state of the science', rather than to present a practical 'how-to-do' for conducting immunotoxicity studies.

There are a number of important concepts that need to be reviewed. First, it is important to understand something about the immune system. Unlike other organ systems, the immune system has the unique quality of not being confined to a single site within the body. The immune system consists of all those physiological processes including cells (i.e., B- and T-lymphocytes, macrophages, natural killer (NK) cells, other inflammatory cells, etc.), organs (i.e., spleen, thymus, bone marrow, lymph nodes, blood, etc.) and soluble mediators (i.e., immunoglobulins, complement, lymphokines, cytokines, etc.) that enable an animal (i.e., the host) to recognize materials as foreign to itself, and to neutralize, eliminate or metabolize them, with or without injury to its own tissue.

Key to this concept = SELF versus NON-SELF

Examples of "**SELF**" are all of the tissues, organs and cells of the body. Examples of "**NON-SELF**" are a variety of opportunistic pathogens, including bacteria and viruses, and transformed cells or tissues (i.e., tumors). If the immune system fails to recognize as non-self an infectious entity, or the neoantigens expressed by a newly arisen tumor, then the host is in danger of succumbing to the unopposed invasion. This aspect of the immune system is the reason why it is often made synonymous with "host defense". Alternatively, if some integral bodily tissue, organ or cell is not identified as self, then the immune system is capable of turning its considerable destructive capabilities against that tissue, organ or cell, and an autoimmune disease may result. The costs of making mistakes in the recognition of self versus non-self can be quite high. The fact that

mistakes can occur in both directions has led some investigators to consider immunotoxicology as a "continuum" (Burns-Naas et al., 2001; Holsapple and Kaminski, 1998). Because of time and space limitations, this lecture will only focus on the suppressive side of the continuum.

Second, it is important to understand a definition of "immunotoxicology". For the purpose of this lecture, "**immunotoxicology**" will be defined as the study of adverse effects on the immune system resulting from exposure to drugs, environmental chemicals or other biological materials.

Key to this definition = "ADVERSE EFFECT"

The need to determine whether an effect is adverse is what differentiates toxicology from other branches of the biomedical science. The term, "**adverse**" has been defined in classical toxicology as "*the undesired effects of a xenobiotic, which are deleterious.*" (Eaton and Klaassen, 1996). Consequently, immunotoxicology is not merely the demonstration of treatment-related changes in a component of the immune system. Not all treatment-related changes are adverse; some may be beneficial; some may be indifferent; and some are of unknown or uncertain consequences. It is inappropriate to declare an effect to be adverse simply because an adverse consequence cannot be ruled out. The long-term credibility of any scientific discipline depends on involved scientists being forthcoming when there is uncertainty, or when an effect has no known adverse consequence. The evolution of the science of immunotoxicology has been based in large part on the design and validation of critical experimental approaches to address the question: Is an observed effect adverse, or is it not? Studies by Luster and co-workers (1988, 1992, 1993) demonstrated that the most predictive indicators of immunotoxicity (i.e., as manifested by immunosuppression) were functional parameters. As such, for the purposes of this lecture, an adverse effect in immunotoxicology will be defined as a chemical-induced suppression in the ability to perform an immune function.

If one considers the fact that immunotoxicology is still evolving, and accepts the definition of "**evolution**" as a " . . . *gradual process of change* . . . ", then it is important to accept the fact that any presentation of the "state of the science" will represent only a "snapshot in time". One such "snapshot in time" for immunotoxicology was provided by the Environmental Defense Fund (EDF) in 1997 for a project entitled, "Toxic Ignorance".

Through the Freedom of Information Act (FOIA), the EDF requested information from all of the major chemical companies in the U.S. regarding the known toxicity for the major chemicals in commerce. While their findings indicated that >80% of these chemicals had genetic toxicity results and >60% of these chemicals had developmental toxicity results, <15% of these chemicals had any immunotoxicity results. It is important to emphasize that the results being analyzed by the EDF were from guideline studies to support the registration of products, that the EPA only published their immunotoxicity guidelines (OPPTS 870.7800) in final form in 1999, and that tests to address immunosuppression are still not required by the EPA. As such, it is highly probable that the number of chemicals for which there are immunotoxicity results that meet the EDF criteria has not greatly increased since 1997. It is also important to emphasize that in 1997, there were hundreds of papers in the peer-reviewed literature describing the immunotoxicity of chemicals, but that the great majority of these papers came from university investigators who do not apply GLP and were based on high dose, acute exposure studies. The significance of the latter points are that industry labs must submit GLP studies to support the registration of their products, that the current draft EPA immunotoxicity guidelines specify exposure for 28 or 90 days, and that high dose, acute exposure studies introduce a high potential for indirect effects on the immune system due to the nonspecific effects associated with stress. Luster et al. (1992) obviously appreciated the role of "dose" because they emphasized the importance of a study design to identify potential immunotoxicants using dose levels that did not induce overt toxicity (i.e., body weight changes >10% or gross pathological changes). There is no doubt that the number of immunotoxicity papers in the peer-reviewed literature has most definitely increased substantially since 1997, and that there have been tremendous advances in our understanding of the mechanistic basis for an approach to address the immunotoxic potential of a chemical.

This lecture will address the mechanistic foundation for immunotoxicology by describing the basis for a model to measure immunosuppression, and by providing an overview of the database generated by Luster and co-workers for the National Toxicology Program (NTP; 1988, 1992, 1993). As previously described, the role of the immune system may be succinctly stated as the preservation of integrity (Karras and

Holsapple, 1996), and immunity can be considered a series of delicately balanced, complex, multicellular and tightly regulated physiologic mechanisms designed to identify self versus non-self (Burns-Naas et al., 2001). Indeed, the tightly orchestrated cellular cooperation that is necessary for the generation of a primary antibody response to a T-cell dependent antigen provides the basis for the sensitivity of this parameter. Not only are B-cells necessary as the ultimate effector, i.e., antibody-forming cells (AFC), but T-cells, i.e., both T-helpers and T-suppressors, are required to regulate the response, and antigen-presenting cells, i.e., macrophages or dendritic cells, are required to get the whole process started. Deficits in the numbers or functional capabilities of any of these cells can be manifested as a suppression of the ultimate antibody response. Methods are available to quantify the number of AFC, and techniques using different types of antigenic stimuli have been demonstrated to further characterize the role of the various types of immune cells in an antibody response (Holsapple, 1985). For example, studies have shown that exposure to carbon tetrachloride selectively suppressed T-helper cells (Kaminski et al., 1989; Delaney and Kaminski, 1993); treatment with casein targeted macrophage antigen-presenting cells (Kaminski and Holsapple, 1987 and 1992); and exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin primarily affected B-lymphocytes (Dooley and Holsapple, 1986; 1988). Perhaps the single most significant observation from this series of studies was that even though these three chemicals selectively targeted different cell types, they all caused a robust suppression of the primary antibody response to a T-cell dependent antigen, such as sheep red blood cells (SRBC).

A multi-lab effort under the primary direction of Dr. Michael Luster for the NTP represents the largest database in immunotoxicology, and is, in fact, one of the largest and most systematic investigations of any target organ (1988, 1992, 1993). A total of 51 chemicals were studied in the original analysis and included a multitude of chemical classes: catalysts, solvents, dyes, lubricants, pesticides, disinfectants, drugs, food additives, natural products, etc. It is important to emphasize that the chemicals selected for study by the NTP had to be nominated because of a suspicion that they would target the immune system. The NTP studies also compared a number of immune endpoints including several functional parameters, host resistance models, organ weights for select immune organs and leukocyte counts & differentials. Luster et al. (1992) considered a

chemical as causing an adverse effect on the immune system if it caused a dose-related suppression of any one parameter or a significant suppression of two or more parameters only at the high dose. In either event, the parameter or parameters affected would be identified as being capable of detecting the immunotoxicity for that chemical. The results of the original analysis are generally presented in the context of the "concordance pyramid", an analysis that considers both sensitivity and specificity. These results confirmed systematically what had been suspected for many years: no single parameter could be used to predict the potential for immunotoxicity. Some of the best assays in terms of concordance included the following: flow cytometric analysis of lymphocyte surface markers (83%), the primary antibody (Ab) response to SRBC (78%), natural killer (NK) cell activity (69%) and thymus/body weight ratio (68%). The analysis by Luster and co-workers indicated that combinations of two tests improved the concordance, as evidenced by the following examples: Ab response + NK activity (94%), Ab response + thymus weight (92%), Ab response + flow (91%) flow + NK activity (90%), flow + thymus weight (90%). A configuration consisting of three tests was required in order to achieve 100% concordance, as evidenced by the following examples: Ab response + thymus weight + NK activity, Ab response + thymus weight + delayed hypersensitivity response (DHR), Ab response + NK activity + DHR, Ab response + flow + DHR.

The studies by Luster et al. clearly identified the primary Ab response to SRBC as one of the most sensitive and specific endpoints in predicting immunotoxicity, in agreement with the discussion above. Because the NTP studies included some attention to intra-laboratory comparisons (Luster et al., 1988), this endpoint is considered by many to be validated. Although the issue of validation should not be taken lightly, there is little doubt that the anti-SRBC Ab response has been the focus of study in immunotoxicology by more laboratories than any other immune parameter. The selection of this endpoint can be attributed to a number of factors. First, it is an *in vivo* correlate for an immune response that is dependent on the cooperation of multiple cell types, as described above. Second, it can be measured in several species, and it has been particularly well studied in both rats and mice. Third, depending on the antigen used to elicit the response, it can be measured either as the number of AFC, as discussed above, or as a serum antibody titer,

as measured by an enzyme-linked immuno-sorbent assay (ELISA), which provides some technical flexibility.

A Science Advisory Board (SAB) for the EPA obviously recognized the sensitivity of the primary Ab response in 1998, as it was one of the assays recommended for inclusion in the immunotoxicity test guidelines. The members of the SAB also recognized that, in spite of several attractive features, the concordance of the Ab response alone was only 78%, which could mean that 1/5 chemicals tested with this assay alone would be missed. As such, the SAB recommended two endpoints in addition to the Ab response: flow cytometric analysis of lymphocyte subsets and the NK cell assay.

SAB Recommendation = Ab response + Flow + NK activity

During the public comment period, this approach was criticized as requiring at least 2 (and perhaps as many as 3) additional groups of animals. As emphasized above, the results of the EDF project emphasized that from a regulatory context, immunotoxicology is still a new discipline, and it is likely that data gaps on immunotoxic potential will exist for many chemicals. The specific configuration of the tests to address this potential will impact on how the gaps are filled, i.e., the length of time and the number of animals required to fill in the immunotoxicity data gaps will be contingent on the testing configuration. For example, 100 chemicals could be characterized with one test in about the same time and expense as 50 chemicals could be characterized with two tests, and 33 chemicals could be characterized with three tests. There is little doubt that whatever testing configuration was recommended, it would have to consider the social pressures that advocate a reduction in the number of animals currently being used in biomedical research. Of course, it is difficult to justify the use of fewer animals if this decision were to diminish the scientific rationale in the context of the confidence that the testing configuration achieved the ultimate goal of predicting immunotoxic potential.

During the public comment period, a number of additional scientific concerns were raised. While the NTP database clearly indicated that flow cytometric analysis was the single "best" parameter with a concordance of 83%, it was never emphasized that a chemical was judged to be "positive" with this approach if it affected either B-cell numbers, total T-cell numbers, T-helper cell numbers, or T-suppressor cell numbers. No other parameter was given so many chances to be labeled as the basis for an adverse

effect. More importantly, it was emphasized in a workshop on the use of flow cytometric analysis of lymphocyte subsets that this approach is not an assessment of immune function, does not always correlate with changes in functional parameters and was not an appropriate "first tier" assay (Sandler et al., 2001). While the NTP database also indicated that the NK cell assay was a predictive endpoint with a concordance of 69%, a number of technical issues were also raised regarding the use of this parameter. Many questions remain regarding the appropriate label for tumor target cells (i.e., ^{51}Cr versus a non-radioisotopic method), acceptable levels of spontaneous release, the "unit" of effect (i.e., % killing at one effector:target ratio versus a lytic unit that is calculated using all ratios), and the definition of an effect (i.e., biological versus statistical significance). Concerns were also raised over the inherent variability associated with control responses in this assay. For example, in the studies by Luster et al. (1988), three laboratories evaluated a number of chemicals for immunotoxicity, and reported baseline NK activity ranging from 3 - 20%. It is important to emphasize that this wide range of "control" values occurred despite the facts that these were three highly qualified laboratories, and that the same inbred strain of mice, common reagents and a common protocol were used. It was also emphasized that the database on the NK activity in rats is even less compelling than that which has been generated in mice. Finally, it was argued that the inclusion of the NK assay was not consistent with the NTP database in that no chemicals were found to be exclusively positive in this test, and no chemicals that were negative in the Ab response were positive in the NK cell assay. The bottom line is that there is not a sufficient database to warrant the inclusion of the NK cell assay as a required test for immunotoxicity.

One final argument was presented during the public comment period to the EPA SAB. The concordance of the Ab response could be improved to >90% if it was coupled with any of the following endpoints: thymus weight, NK activity or flow cytometric analysis. As noted above, the inclusion of either of the latter two endpoints would mean additional animals. In contrast, thymus weight is an endpoint that is either already being routinely measured in studies done in the chemical industry, or could easily be integrated into existing studies. In this regard, the studies by Ladics et al. (1995) are important because they showed that the Ab response could be measured in a standard toxicology

study indicating that the Ab response and thymus weights could be measured in one group of animals.

It is clear that the combination of Ab response plus thymus weights would require fewer animals than either of the other combinations, and that, based on the NTP studies would predict immunotoxic potential with equal confidence to the other combinations. However, it must still be recognized that the combination of Ab response + thymus weight had a concordance of 92%, which could mean that 1/10 chemicals tested in this configuration would be missed. Therefore, it is important to look for ways to improve the concordance of the Ab response + thymus weight. As noted above, the studies by Luster et al. (1992) indicated that one would have to add either the NK assay or the DHR to this combination to achieve 100% concordance. Either option would add significantly to the cost and complexity of the immunotoxicity testing configuration. While thymus weight was projected in the NTP studies as a highly predictive endpoint with a concordance of 68%, it is widely recognized to be far less informative than a histopathological analysis of the thymus. Histopathology was not emphasized in the original NTP studies; but it is important to note that the NTP is currently in the process of re-analyzing tissues from the earlier studies and/or engaging in a multi-lab analysis of new tissues from more recent studies with some of the same chemicals that were used in the original investigations, with the goal being to improve the quality of histopathology in their analysis. Because histopathology of immune organs has been previously shown to provide clear indications of the potential for damage to the immune system (Basketter et al., 1994, 1995), there is little doubt that the concordance of a testing strategy based on Ab response + thymus histopathology would be > the concordance of Ab response + thymus weight (i.e., 92%). Moreover, it is important to emphasize that from a practical perspective, the histopathological analysis conducted in an industry lab would not be limited to thymus, and would also include spleen, lymph nodes and other immune organs, which should further improve the concordance.

The final draft of the EPA Immunotoxicity Test Guidelines (OPPTS 870.7800) was published in January, 1999. The primary Ab response to SRBC was the only required endpoint, and could be assessed in either rats or mice, after either 28 or 90 days of exposure, by either the AFC assay or the ELISA. The other two endpoints originally

recommended by the EPA SAB were relegated to "optional" tests that could be triggered based on the outcome of the analysis of the Ab response. Specifically, if the Ab response was "positive" (i.e., the test chemical caused suppression), phenotypic analysis by flow could be done, and if the Ab response was "negative" (i.e., the test chemical had no effect on the Ab response), the NK assay could be done. As recently emphasized (Burns-Naas et al., 2001) this testing configuration does not represent a comprehensive assessment of immune function but is intended to complement assessments made in routine toxicity testing such as hematology, lymphoid organ weights and histopathology. The identification of phenotypic analysis as a "second tier" assay is consistent with a recent workshop that noted the tremendous potential of this approach as a mechanistic tool (Sandler et al., 2001). However, the identification of the NK assay as an approach to detect chemicals that are devoid of activity in the Ab response is not based on a strong scientific database, as noted above. Ultimately, the biggest concern with the proposed guidelines is the fact the specific criteria that will be used to trigger the "optional" tests have never been identified.

In conclusion, the state of the science of immunotoxicology has evolved, especially over the last ten years. There are immune tests, such as the primary Ab response, for which there is a solid mechanistic foundation. The predictive value of this test was clearly demonstrated in the NTP database, and our confidence in a testing configuration based on the Ab response is increased when this test is interpreted in the context of routine toxicity studies done according to GLP. However, there are still challenges ahead that will require the continued evolution of immunotoxicology. One of the greatest challenges that was identified in a workshop several years ago (Neumann et al., 1995) is how to interpret the significance of minor or moderate immunotoxic effects observed in rodents as it relates to risk assessment to humans.

References

Basketter, D.A., Bremmer, J.N., Kammuller, M.E., Kawabata, T.T., Kimber, I., Loveless, S.E., Magdapal, T.H.M., Stringer, D.A., and Vohr, H.W. (1994). Pathology considerations for, and subsequent risk assessment of, chemicals identified as immunosuppressive in routine toxicology. Fd. Chem. Toxic. 32:289-296.

Basketter, D.A., Bremmer, J.N., Buckley, P., Kammuller, M.E., Kawabata, T.T., Kimber, I., Loveless, S.E., Magda, S., Stringer, D.A. and Vohr, H.W. (1995). The identification of chemicals with sensitizing or immunosuppressive properties in routine toxicology. Fd. Chem. Toxic. 33:239-243.

Burns-Naas, L.A., Meade, B.J. and Munson, A.E. (2001). Toxic response of the immune system. In: TOXICOLOGY: The Basic Science of Poisons. (McGraw-Hill). Pp. 419-470.

Delaney, B. and Kaminski, N.E. (1993). Induction of serum borne immunomodulatory factors by carbon tetrachloride. I. Carbon tetrachloride-induced suppression of T_H-cell activity is mediated a serum borne factor(s). Toxicology 85:67-84.

Dooley, R.K. and Holsapple, M.P. (1986). Cellular targets responsible for 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD)-induced immunosuppression. J. Leuk. Biol. 40:309-310.

Dooley, R.K. and Holsapple, M.P. (1988). Elucidation of cellular targets responsible for tetrachlorodibenzo-p-dioxin (TCDD)-induced suppression of the antibody response. 1. Role of the B-lymphocyte. Immunopharmacology 16:167-180.

Eaton, D.L. and Klaassen, C.D. (1996). Principles of toxicology. In: TOXICOLOGY: The Basic Science of Poisons. (McGraw-Hill). Pp. 13-34.

Holsapple, M.P. (1995). The plaque forming cell (PFC) response in immunotoxicology: An approach to monitoring the primary effector function of B-lymphocytes. In: Modern Methods of Immunotoxicology (Vol. 1). (Wiley-Liss). Pp. 71-108.

Holsapple, M.P. and Kaminski, N.E. (1998). Immune System. In: Encyclopedia of TOXICOLOGY (Vol. 2). (Academic Press). Pp. 114-139.

Kaminski, N.E. and Holsapple, M.P. (1987). Inhibition of macrophage accessory cell function in casein-treated B6C3F1 mice. J. Immunol. 139:1804-1810.

Kaminski, N.E., Jordan, S.D. and Holsapple, M.P. (1989). Suppression of humoral and cell-mediated immune responses by carbon tetrachloride. Fundam. Appl. Toxicol. 12:117-128.

Kaminski, N.E. and Holsapple, M.P. (1992). A functional characterization of macrophage alterations in casein-treated mice. Immunopharmacology 24:229-240.

Karras, J.G. and Holsapple, M.P. (1996). Structure and function of the immune system. In: Experimental Immunotoxicology. (CRC Press). Pp. 3-12.

Ladics, G.S., Smith, C., Heaps, K., Elliot, G.S., Slone, T.W. and Loveless, S.E. 1995. Possible incorporation of an immunotoxicological functional assay for assessing humoral immunity for hazard identification purposes in rats on standard toxicology study. Toxicology 96:225-238.

Luster, M.I., Munson, A.E., Thomas, P.T., Holsapple, M.P., Fenters, J.D., White, K.L., Jr., Lauer, L.D., Germolec, D.R., Rosenthal, G.J. and Dean, J.H. (1988). Development of a testing battery to assess chemical-induced immunotoxicity: National Toxicology Program's guidelines for immunotoxicity evaluation in mice. Fundam. Appl. Toxicol. 10:2-19.

Luster, M.I., Portier, C., Pait, D.G., White, K.L., Jr., Gennings, C., Munson, A.E. and Roenthal, G.J. (1992). Risk assessment in immunotoxicology. I. Sensitivity and predictability of immune tests. Fundam. Appl. Toxicol. 18:200-210.

Luster, M.I., Portier, C., Pait, D.G., White, K.L., Jr., Rosenthal, G.J., Germolec, D.R., Corsini, E., Blaylock, B.L., Pollock, P., Kouchi, Y., Craig, W., Munson, A.E. and Comment, C.E. (1993). Risk assessment in immunotoxicology. II. Relationship between immune function and host resistance tests. Fundam. Appl. Toxicol. 21:71-82.

Neumann, D. and Immunotoxicology Technical Committee (1995). ILSI Health and Environmental Sciences Institute: Immunotoxicity testing and risk assessment: Summary of a 1994 workshop. Fd. Chem. Toxic. 33:887-894.

Sandler, D. and Immunotoxicology Technical Committee (2001). ILSI Health and Environmental Sciences Institute: Applications of flow cytometry to immunoxicity testing: Summary of a workshop. Toxicology 169:39-48.