The Road to Embryologically Based Dose–Response Models

Robert J. Kavlock and R. Woodrow Setzer

National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina

The goal of researchers working in the area of developmental toxicology is to prevent adverse reproductive outcomes (early pregnancy loss, birth defects, reduced birth weight, and altered functional development) in humans due to exposures to environmental contaminants, therapeutic drugs, and other factors. To best achieve that goal, it is important that relevant information be gathered and assimilated in the risk assessment process. One of the major challenges of improved risk assessment is to better use all pertinent biological and mechanistic information. This may be done qualitatively (e.g., demonstrating that the experimental model is not appropriate for extrapolation purposes); semiquantitatively (using information to reduce the degree of uncertainty present under default extrapolation procedures), or quantitatively (formally describing the relationships between exposure and adverse outcome in mathematical forms, including components that directly reflect individual steps in the overall progression of toxicity). In this paper we review the recent advances in the risk assessment process for developmental toxicants and hypothesize on future directions that may revolutionize our thinking in this area. The road to these changes sometimes appears to be a well-mapped course on a relatively smooth surface; at other times the path is bumpy and obscure, while at still other times it is only a wish in the eye of the engineer to cross an uncharted and rugged environment. — Environ Health Perspect 104(Suppl 1):107–121 (1996)

Key words: risk assessment, RfD, developmental toxicity, teratogencity, benchmark dose, biologically based dose–response models, BBDR, homeobox, skeletal development, limb development, pattern formation, mathematical models

During the last 5 years, significant changes in the risk assessment process for noncancer health effects of environmental contaminants have begun to appear. The first of these changes is the development and use of statistically based dose–response models (the benchmark dose approach) that better utilize data derived from existing testing approaches. Accompanying this change is the greater emphasis on understanding and using mechanistic information to yield more accurate, reliable, and less uncertain risk assessments. The next stage in the evolution of risk assessment will be the use of biologically based dose–response (BBDR) models to build factors related to the underlying kinetic, biochemical, or physiological processes, which may be perturbed by a toxicant, into the statistically based models. Such models are now emerging from several research laboratories. The introduction of quantitative models and the incorporation of biological information into them has pointed to the need for even more sophisticated modifications, which we term embryologically based dose–response (EBDR) models. Because these models are based upon the understanding of normal morphogenesis, they represent a quantum leap in our thinking, but their complexity presents daunting challenges both to the developmental biologist and the developmental toxicologist. However, the remarkable progress in the understanding of mammalian embryonic development at the molecular level that has occurred over the last decade should eventually enable these as yet hypothetical models to be brought into use.

A firm understanding of the mechanisms of normal development is required to adequately characterize mechanisms of abnormal development. Indeed, the paucity of complete descriptions of mechanisms of chemically induced dysmorphogenesis is in large part based on our poor understanding of normal developmental processes. For example, without an understanding of the forces that control the outgrowth and differentiation of the limb bud, how can we understand the formation of limb deformities? Advances in understanding morphogenesis on the molecular and biochemical level, for the first time, are providing the knowledge base necessary for developmental toxicologists to truly understand the mechanisms by which chemicals disrupt embryogenesis. Clearly, we are still far from having models of normal morphogenesis commonly available in the toolbox of the developmental toxicologist and risk assessor, but one day we may witness a revolutionary change not only in how we evaluate developmental toxicity in animal models but also in how toxicity is extrapolated to the human population.

Basic Elements of Dose–Response Assessment

For the purposes of this review, dose–response assessment can be viewed as three critical steps: identification of the effect (and related exposure level) of most concern; a characterization of the uncertainty

Manuscript received 14 August 1995; manuscript accepted 1 November 1995.

We thank Lynn Connelly for compiling the table of phenotypes from the null hox mutants and Phil Hartig, John Rogers, Sid Hunter, and John Creason for discussions and comments on the manuscript.

This manuscript has been reviewed by the National Health and Environmental Effects Research Laboratory and approved for publication. Mention of trade names or commercial products does not constitute endorsement.

Address correspondence to Dr. Robert J. Kavlock, National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711. Telephone: (919) 541-2326. Fax: (919) 541-1499. E-mail: kavlock@herl45.herl.epa.gov

Abbreviations used: BBDR, biologically based dose–respose model; EBDR, embryologically based dose–response model; NOAEL, no observed adverse effect level (synonymous in this paper with NOSTASOT); NOSTASOT, the highest experimental dose, which, upon dropping all higher doses, no longer results in a positive trend test; RfD, reference dose; RfC, reference concentration; BMD, the benchmark dose or the lower confidence limit of the dose estimated to be MLE; MLE, the maximum likelihood estimate of the dose inducing a 5% added risk of effect as determined by a log–logistic dose–response model that incorporates litter size and intralitter correlations; Q, quantal; C, continuous; QNOAEL, the NOAEL for quantal (litter based) response measures in developmental toxicity assays; CNOAEL, the NOAEL for continuous (fetal-based) response measures in developmental toxicity assays; BME, benchmark effect; 2,4,5-T, trichlorophenoxyacetic acid; 5-FU, 5-fluorouracil; GD, gestation day; TS, thymidylate synthetase; AER, apical epidermal ridge; PZ, progress zone; FGF, fibroblast growth factor; ZPA, zone of polarizing ability.

present in the database; and an estimate of the exposure level presumed to be free of risk to the human conceptus. In the first step, data from exposed experimental species, as well as any epidemiological information, is examined for the highest dose level that is without a significant adverse effect. This dose level is established by a combination of statistical analysis and expert opinion and is generally referred to as the NOAEL (the no observed adverse effect level). It is important not to confuse the concept of NOAEL with that of threshold for biological effects, as the former merely reflects the statistical power of an experiment to see an effect when in fact one does exist. Various regulatory agencies have provided guidelines for the design, conduct, and interpretation of such hazard identification studies for developmental toxicity (*1*). The lowest NOAEL in the database on a particular chemical is termed the critical NOAEL. In the second step, the adequacy, relevance, and uncertainties in extrapolating the NOAEL from the experimental species to the target species are estimated. Minimally, these extrapolations consider the sensitivity of the human, relative to test species, and the presence of potentially sensitive subpopulations. In the default situation, uncertainty factors of 10 are used to cover both sources of uncertainty. Other uncertainty or modifying factors may be applied to account for a lack of identification of a NOAEL, an incomplete database, or an expert opinion regarding the probability of risk. In the final step, the critical NOAEL is divided by the product of the uncertainty factors, as well as any expert-derived modifying factors (MF) to obtain the reference dose [RfD] (or reference concentration (RfC) for an inhaled chemical).

$$\mathrm{RfD} = \frac{\mathrm{NOAEL}_{\text{critical effect}}}{\mathrm{UF}_{\text{interspecies}} \times \mathrm{UF}_{\text{intraspecies}} \times \mathrm{MF}}$$

Lifetime exposures below the RfD or RfC are believed to be without appreciable risk to humans (*1–4*).

Qualitative and Semiquantitative Approaches

Perhaps the easiest and most straightforward approach to incorporating mechanisms into dose–response assessment occurs when it can be demonstrated that the animal model yields results that are not extrapolatable (a process that can be called qualitative nonextrapolation). While this may be pertinent only on rare occasions, the impact is always profound. Excellent examples of this have been derived by research into the mechanisms of carcinogenesis, where we now understand that saccharin-induced bladder tumors in rodents and α_2-microglobulin-induced kidney tumors in male rats have no homologous counterpart in humans. Likewise, if a metabolite is found to be the proximate toxicant and that metabolite is not formed in humans, then the subsequent effect would not be expected to occur either. In developmental toxicity, examples include the effects of Gram-negative antibiotics in rabbits in which the marked effect of the chemical class on the gut microflora causes a nutritional deficiency that secondarily produces effects on the offspring (*5*) and the effect of diflunisal-induced axial skeletal defects in rabbits that are due to a species-specific maternal hemolytic anemia (*6*).

Another approach to incorporating biological understanding into the dose–response and risk assessment activities is to carefully examine the quality, consistency, and adequacy of the database in light of the default assumptions regarding the presumed sensitivity of humans relative to the experimental species and to the existence of susceptible subpopulations. If the database provides sufficient evidence, it is then possible to reduce the magnitude of the uncertainty factors to reflect the level of understanding of interspecies and intraspecies differences. Such an approach is exemplified by the Institute for Evaluating Health Risks Evaluative Process (*4*) in the assessment of the reproductive and development effects of lithium (*7*). The human and experimental evidence were judged sufficient to indicate that lithium causes developmental toxicity in the therapeutic range, but it fell short of indicating what the presumed safe level of exposure would be; extrapolation therefore was necessary. An expert review committee concluded that the uncertainty factor of 10 for interspecies differences could be reduced to $10^{0.5}$ on the basis of the facts that *a*) Li^{+2} is the active toxicant; *b*) there was a linear relationship between lithium exposure and plasma levels in both humans and experimental animals; *c*) adverse effects appear to occur at similar lithium levels in humans and animals; and *d*) the systemic target organs for lithium toxicity are similar in humans and animals. In addition, the intraspecies uncertainty factor was also reduced to $10^{0.5}$ on the basis that the use of serum concentrations as a measure of delivered dose minimizes interindividual differences in absorption, hence accounting for some of the differences within a population. The aggregate uncertainty factor used was therefore $10^{0.5}$ times $10^{0.5}$, or 10. This effort is one of the first coordinated attempts to bring independent experts together for the specific purpose of providing the best scientific determination of risk of adverse reproductive outcomes (similar in nature to the effort of the International Agency for Research on Cancer for carcinogenesis); this effort also demonstrates the types of decisions that such an informed group can make regarding the magnitude of the uncertainties present in a typical example.

The Benchmark Dose Approach

Another avenue to improve the dose–response component of the risk assessment process is to better use data generated from standardized testing procedures, independent of knowledge of toxicokinetic or toxicodynamic factors that may be used to adjust the magnitude of the uncertainty factors. Reliance on the NOAEL as the entrance point into the extrapolation process for noncancer effects has been the subject of much criticism (*1*). The most significant criticism has arisen from the fact that the procedure to obtain the NOAEL fails to encourage better experimental design. In fact, it even works actively to discourage such efforts: experiments with more dose groups and more subjects per group can only result in lower NOAELs because more statistical power is focused on between-dose-group comparisons. Thus, chemical manufacturers are effectively discouraged from submitting better toxicological data than the minimum required by regulatory agencies. Other criticisms include the need to repeat experiments that fail to demonstrate a NOAEL; the approach ignores the shape and variability of the dose–response curve; and NOAELs can represent considerable (and inconsistent) risk levels (*8,9*).

Many of these criticisms have been addressed by application of statistically based dose–response models in the benchmark dose (BMD) approach (*10*). In the BMD approach, a particular effect level is chosen and the dose inducing that response is calculated using a statistical model (Figure 1). The BMD is then defined as the lower 95% confidence interval on that dose level. In principle, the response level is chosen near the low end of the observable range so no extrapolation is necessary (*11*). The use of a dose–response function brings

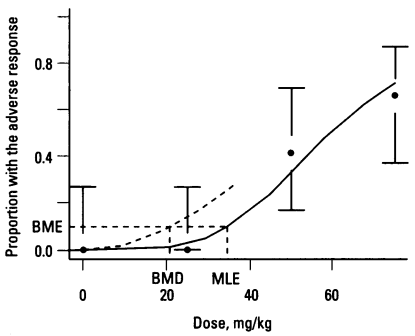

Figure 1. Benchmark dose calculation. In this idealized graph, the experimental data points are depicted by the symbols, the smooth line is the model fit to the data, and the dashed line is the lower confidence limit on dose for a given response level. The BME is shown here as an extra 5% risk, and the BMD is the dose that corresponds to the intersection of the BME and the confidence interval.

data from all experimental doses into use, and the use of the lower confidence interval on the dose estimate for a particular risk allows the experimental variability to enter into the output. BMDs from different end points or different studies would therefore be based on more similar response levels than occurs with the NOAEL.

To evaluate the utility of the BMD to standard developmental toxicity test data, a database of 246 studies was analyzed (*12–15*). These studies used two dose–response models applicable to any toxicological end point (the quantal Weibull model and the continuous power model), as well as three models (termed the RVR, NCTR, and LOG models) that incorporate aspects specific to developmental toxicity data (e.g., litter-size effects and intralitter correlations). These studies also examined quantal (Q) end points (the presence or absence of litters with at least one dead or malformed implant) and continuous (C) end points (the mean litter incidence of affected implants and fetal weight). BMDs for various added risk levels (1, 5, and 10%) were estimated from a variety of model formulations (e.g., the presence of a threshold or litter-size parameter). For incidence data, a total of 1,825 end point-specific BMDs and corresponding NOAELs were determined. For fetal weight, comparisons were based on a subset of studies for which individual fetal weight data were available, and only the continuous power model and the LOG model (with the litter size but without the threshold parameter) were used. To calculate a benchmark dose for reductions in fetal weight, it was first necessary to define what level of effect should be used in the assessment. Therefore, in a preliminary analysis, 18 different definitions of reduced fetal weight were considered in establishing the benchmark effect (BME) level that was similar in magnitude, on average, to traditionally determined NOAELs. Six BMEs for reduced fetal weight were used in the full analysis. These included reductions in the mean fetal weight by 5%, 0.5 standard deviations (SD), or 2 standard errors of the mean (SEM); a reduction in mean fetal weight to the 25th percentile of the control mean; and a 5 or 10% increase in incidence of fetuses weighing less than the 5th or 10th percentile, respectively, of the control litter mean. The NOAEL for reduced fetal weight was less than the highest experimental dose level in 85 of the 173 studies in this subset. χ^2 tests were used to assess goodness of fit while the magnitude of the log-likelihood estimates were used to compare the influence of optional model parameters. BMDs were then compared with traditionally determined NOAELs.

In the database, QNOAELs (the NOAEL for an end point based upon whether a litter contained an affected implant) were similar in magnitude to CNOAELs (the NOAEL for an end point based upon the mean incidence of affected implants within litters). Both generic and developmental-specific models provided acceptable fits to the data from these standard developmental toxicity bioassays. For the generic models, goodness of fit tests were rejected in less than 4% of all analyses. A very low frequency of nonconvergence (4/1,825) occurred, which seemed to coincide with dose–response patterns in which the response at low doses was higher than that at higher dose levels. For the developmental-specific models, incorporation of the litter size but not the threshold parameter marginally improved model fit, and the LOG model was slightly better than the RVR and NCTR models in terms of model fit (probably the result of its more flexible handling of the litter-size parameter).

In nearly every comparison, the median ratios of benchmarks to NOAELs were closer to unity than the means, suggesting the presence of non-normal distributions. In comparisons based upon the quantal approach, the best match on average between NOAELs and the BMDS was found for a $QMBD_{10}$ (a 10% added risk from the quantal Weibull model). The median benchmark-to-NOAEL ratio at this risk level was 0.5, and 88% of the BMDs were within a factor of 5 of the NOAEL. When comparisons were based upon more continuous measures of response (the mean litter incidence), the best matches were found between the NOAEL and a $CBMD_{05}$ (a 5% added risk from the continuous power model), and a BMD_{05} from any of the three developmental-specific models. For example, the median CBMD-to-NOAEL ratio was 1.04, and 95% of the benchmarks were within a factor of 5 of the NOAEL; and only 9/486 (1.85%) of the comparisons differed by a factor of 10. All six operational definitions of reduced fetal weight listed above provided BMDs that were similar, on average to the NOAELs. The median benchmark to NOAEL ratios ranged from a low of 0.9 for a BME of a 2 SEM reduction in the average litter mean to a high of 1.24 for a 5% reduction in the average litter mean. In only 9/76 comparisons (11.8%) did any of the BMDs from the six definitions of a BME for reduced fetal weight differ from the NOAEL by more than a factor of 4. The largest such difference was 18-fold. Two aspects, often in combination, generally contributed to the differences of unusual magnitudes between the BMD_{05} for the reduced fetal weight BMEs and the corresponding NOAELs. The first of these was the use of a study design with wide dose spacing and the second was the presence of a shallow dose–response pattern. In the former instance, a very wide interval between the NOAEL and LOAEL dose would tend to produce NOAELs that might be considered to be artificially low (recall that unlike the benchmark dose, the NOAEL is constrained to be one of the experimental doses). In the latter instance, the shallow slope can make determination of the NOAEL more arbitrary and unstable. Combined, these two aspects can therefore be expected to create greater than normal heterogeneity in the BMD-to-NOAEL ratio. Therefore, close examination of the minority of studies that yielded divergent BMD-to-NOAEL ratios demonstrates a key advantage of the benchmark dose approach. If the BMD always gives us the same information as the NOAEL, there is little compelling reason to adopt a new system. Starting the extrapolation process from a point of consistently determined and comparable risk levels avoids some potentially misleading comparisons of the relative risk of two chemicals.

These comparisons demonstrate the feasibility of applying the benchmark dose methodology to developmental toxicity bioassays and provide convincing evidence

of the risk level (5% on average) associated with traditionally derived NOAELs based upon continuous measures of response. The analysis also points to the conservative nature of dose–response models based upon quantal end points that reduce the data to whether a litter contained at least one affected implant (recall that the QBMD that best matched the QNOAEL was based upon an added risk of 10%, and even in that case the estimate was still conservative relative to the NOAEL). At the 10% risk level, QBMDs were two to three times smaller than the CBMDs, a difference attributable to both lower maximum likelihood estimates (MLE) of the dose corresponding to this risk level and to wider confidence intervals. For example, the median ratio for the $CMLE_{05}/CBMD_{05}$ was 1.6 with an upper 99th percentile of 4.0, whereas the median ratio of the $QMLE_{05}/QBMD_{05}$ was 3.3 with an upper 99th percentile of 15.6. In only 13 of 542 end points with significant quantal and continuous trends, the $QBMD_{05}$ was greater than the $CBMD_{05}$. Such comparisons may have important implications for implementing the BMD approach for other noncancer end points in which data are more like those of the quantal end points that are based upon affected litters and point to the advantages of using individual implant data for developmental toxicity analyses. In any event, there were no major differences in performance or fitness between the generic, continuous, or developmentally specific benchmark models for any of the end point comparisons, suggesting that the choice of model is up to the user, provided that it adequately describes the data. Another advantage of the BMD approach is that it provides a stimulus for considering other dose–response methodologies in the area of noncancer health effects, a situation that appeared extremely remote only a few years ago.

Second Generation Models

The discussion to this point has been limited to assessing each end point of developmental toxicity individually. Several groups are now working on models that are capable of providing risk estimates for multiple adverse outcomes. End points of developmental toxicity nominally recorded in bioassays include the viability of an implant and the morphological status and growth of surviving implants. Emerging models can account for correlations between outcomes and can represent the overall probability of yielding a normal birth outcome. For example, Catalano et al. (*16*) presented a model of the probability of abnormality, which is defined as the probability that an offspring is either dead, malformed, or of low fetal weight. The model can be simply expressed as

$$P(d) = 1 - [1 - P_1(d)]\,[1 - P_2(d)]$$

where $P(d)$ is the overall probability of being normal at dose d, P_1 is the probability of death or resorption, and P_2 is the probability of malformation or low weight conditional on survival. The models were fit using generalized estimating equations (GEE), which are computationally simpler than maximum likelihood methods and also have relaxed distributional assumptions. The probability of fetal death or resorption was initially modeled as a function of dose using a probit model with a power parameter. Next, outcomes among live fetuses were modeled using a two-stage regression approach. The first stage regresses fetal weight as a function of dose with litter size as a covariate and allows for a correlation parameter to characterize the litter effect. Then, the procedure calculates the individual and average litter residuals from the fetal weight model. A cutoff value of 3 SD units beneath the control mean was considered abnormal. Next, a probit model was used to quantify the probability of malformation, with covariates for dose, individual and average weight residuals, and litter size. Finally, all three models were linked to obtain an overall risk of adverse outcome. An important model assumption was that malformation and weight are independent after conditioning on litter size. The multinomial approach should produce more conservative estimates of adverse outcome as a result of its increased power and sensitivity to detect effects among strongly correlated outcomes. Thus, in the analysis of the effects of diethylene glycol dimethyl ether (DYME) on mouse development, the individual estimates of the BMD_{05} for death, malformation, and weight were 152.4, 141.7, and 149.8 mg/kg, respectively, while the full multinomial yielded a combined BMD_{05} of 99.4 mg/kg.

Similarly, Zhu et al. (*17*) examined an extended Dirichlet-multinomial covariance function to estimate jointly the regression parameters in Weibull dose–response models for both embryolethality and fetal malformations as applied to the large-scale study conducted by the National Center for Toxiocological Research on the developmental toxicity of 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) in mice. Here, the fetal malformation rate was determined conditionally on both the number of implants and the number of viable fetuses. Using GEE to estimate the model parameters, the doses associated with 5% increased risk of response (ED_{05}) were: 51.93 mg/kg for cleft palate, 55.55 mg/kg for prenatal death, and 46.51 mg/kg for what they termed overall toxicity. Krewski and Zhu (*18*) later extended the comparison of binomial and trinomial models to 11 datasets and found that when both end points were affected by dose, the ED_{05}s were always lower for overall toxicity and the standard errors of the estimate tended to be smaller. When only one end point was affected by dose, the ED_{05} for overall toxicity approximated that for the affected end point.

As seen in both examples of multinomial approaches to developmental toxicity, there is a gain in sensitivity (conservatism) in estimating the joint probability of response when multiple end points are affected by exposure. As many of the measured end points are intercorrelated, and perhaps even related biologically, there seems to be considerable logic in models that are capable of estimating the overall risk of adverse outcome; there are also computational advantages of these models. However, they have not been extensively evaluated in large numbers of datasets, and the gains in precision generally appear to be relatively small.

Study Design Implications of the BMD Approach

The practical consideration of identifying the highest experimental group that does not differ significantly from the control group has led to study designs for developmental toxicology that generally consist of four dose groups (one control and three treated) of 20 litters each. The high dose is usually targeted to induce mild maternal toxicity, with lower doses set either by progressively halving the higher doses or by other factors, such as the desire to ensure that no adverse maternal or developmental effects are observed at the lowest experimental dose. Given the sample sizes and background response rates, these designs are generally capable of detecting a 3- to 6-fold increase in embryonic death, a 5- to 12-fold increase in malformations, and a 15 to 20% decrease in mean fetal weight (*1*). With the emergence of the BMD approach for dose–response assessment, the premium on identification of the NOAEL is likely to diminish in favor of designs that

yield smaller confidence intervals, and hence higher BMDs, around the benchmark effect level. Given this new consideration, it is important to analyze how elements of study design can influence estimation of the BMD. While there has been considerable effort placed on examining the influence of study design on cancer risk assessment models, the fact that target risk estimates for noncancer effects lie in the observable range (10^0–10^{-1}) as opposed to very low and experimentally nonconfirmable levels for carcinogenic effects (generally 10^{-5}–10^{-6}) provides a different set of issues related to study design. Thus, it is expected that the BMD will be less sensitive to model misspecification, provided that the models are flexible in terms of handling different dose–response patterns (*11*). The most important issue for BMD calculations is how the number of dose groups, their spacing, and sample size affect the precision and accuracy of the risk measure. These aspects were studied by simulating the dose–response effects of 5-fluorouracil (5-FU) on the incidence of malformations and reduced fetal growth noted in a study by Shuey et al. (*19*). The primary findings of that study will be presented in the discussion on biologically based dose models. In the simulation study (*20*), fixed sample size designs were studied: a total of 80 litters were distributed as evenly as possible among four, five, six, seven, or eight dose levels, as well as fixed dose group designs in which 10, 13, 17, or 20 litters per group were assigned to either four or five strategically spaced dose levels. In this review, discussion will be focused on malformations as the outcome variable.

The observed and simulated dose response patterns for 5-FU-induced malformations are depicted in Figure 2A. Note that the response pattern is steep; elevated responses are first observed at 30 mg/kg and increase rapidly to reach nearly 70% by 40 mg/kg. In the original experimental data, the BMD for malformations was 26.5 mg/kg with a NOAEL of 30 mg/kg. Figure 2*B* presents a composite analysis of the results from five different simulated designs. The designs varied from four through eight dose groups, with all dose groups consisting of 10 litters. Fifty permutations were run for each design; benchmarks were calculated using the log–logistic model with litter size and intralitter correlations as optional parameters. NOAELS were determined by the NOSTASOT method (*21*). Several elements can be observed in Figure 2*B*: the proximity (and variability) of the MLE estimations to the true dose level that induced a 5% increase in malformations (the ED_{05} of the simulated population); the width (and variability) of the confidence interval beneath the MLE (the BMD); and the impact on the NOAEL. For example, if the study design contained a dose group at 30 mg/kg, that dose was identified as the NOAEL the majority (72–78%) of the time. The deviation for the true ED_{05} can be gauged from the mean squared error (MSE) of the MLE, while its bias can be gauged from the absolute difference from the true ED_{05}. Regardless of the design, the median MLEs were never far from the ED_{05}s; the largest differences were an overestimation of about 10% for a design with only one dose, with a positive response and an underestimation of about 10% when the design carried two responding dose levels, neither with a response level near the ED_{05} (i.e., there was no dose group at 30 mg/kg). In both of these cases, the confidence interval for the MLE was large, with mean BMD levels as low as 50% of the MLE. The dose-spacing and dose–response patterns for the design that contained 0, 10, 20, and 40 mg/kg are somewhat like those observed with a typical design with an active developmental toxicant. The variability of the MLE was lowest and the magnitude of the BMD highest (e.g., smallest confidence intervals for the MLE) when the design contained two positive dose groups, one of which was near the true ED_{05} (i.e., one dose group at 30 and one at 35 or 40 mg/kg). Somewhat more variable results were obtained when the only positive dose group was near the ED_{05}. This response pattern is typical of what might be considered a weak developmental toxicant.

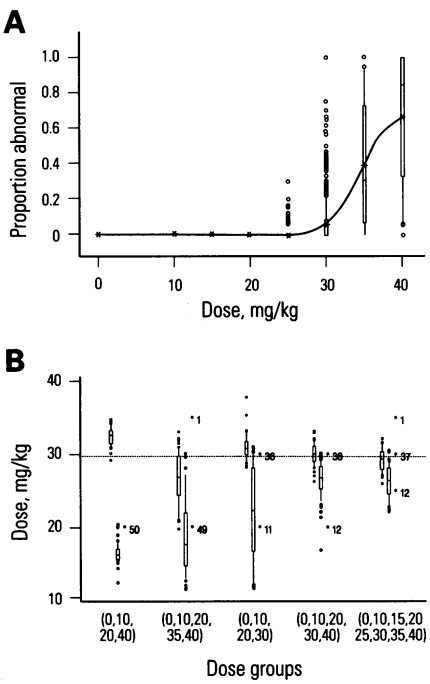

Figure 2. *A*, Box and whisker plot for the dose response for abnormalities derived from 1,000 samplings of the simulated population. The smooth line is drawn through the true population response at each dose level. In the box and whiskers, the box encloses the interquartile range, the lines extending from the box reach the 10th and 90th percentiles, and the line in the box is the median value. Litters with responses in the tail 10th percentiles are indicated by open circles. *B*, Graphical representation of results from simulations of different study designs for abnormalities given 10 litters per dose. For each study design, the box plot on the left is for the MLE, the box plot immediately to the right is for the BMDs, and the number of simulations (50 total) at each dose level giving a particular NOSTASOT is inscribed to the right. Box plots are as described in *A*. Adapted from Kavlock et al. (*20*).

Although the current NOAEL-based strategies for study design in developmental toxicity studies are adequate for use in benchmark dose calculations, the results of this simulation study point to minor improvements that offer increased precision and accuracy of the desired risk estimates. The most important of these factors include the location of a dose group at the low end of the response range and the presence of two nonzero responding dose levels. It is possible that information from pilot dose–range-finding studies could help position dose levels for the definitive dose–response study.

The Biologically Based Approach

In vivo

Thus far we have discussed how to better use data traditionally acquired in the hazard identification phase of risk assessment. We now turn our attention to approaches that require additional knowledge and characterization and, by doing so, offer greater promise of reducing the uncertainties present in extrapolating data from experimental systems to humans. In the first set of examples, the problem is approached from the top, i.e., identifying an effect and putative mechanism, acquiring the necessary information, and constructing a formal quantitative response model. These models, termed pharmacodynamically or biologically based dose–response (BBDR) models, break down the sequence of events—from administration of the chemical to expression of

toxicity (in the case of developmental toxicity these would be an altered phenotype in the offspring)—into individual intervening steps and attempt to quantitatively describe each segment (Figure 3). Thus, to the extent possible and feasible, these models attempt to determine and quantify mechanisms of toxicity and how the relationships may change as a function of dose rate, route of administration, and molecular, biochemical, and physiological differences across species. Much of the early efforts in development of BBDR models has been in the area of carcinogenesis (*22,23*), but attention is now being turned to noncancer end points as well. Gaylor and Razzaghi (*24*) postulated a model to describe the induction of cleft palate by 2,4,5-T (using the same NCTR dataset used above for multinomial benchmark-dose comparisons). To simplify the model, they assumed that the chemical affected only one stage of development through a reduction in cell number, which in turn yielded palatal shelves that were too small to close, hence the cleft. A logistic function was used to describe the growth rate of cells, and an exponential function described the effect of target dose on the growth rate constant. The probability of a normal palate in a treated animal relative to a normal palate in a control animal was assumed to be equal to the ratio of the number of palatal cells present at the critical time of closure. The overall model was expressed as:

$$P(D)-P(0)=[1-P(0)] \times \frac{e^{\beta_0 t' e^{-\alpha D\gamma}}-e^{-\beta_0 t'}}{1-\beta_0 t'}$$

where $P(D)$ is the probability of a cleft palate at dose D, β_0 is the growth rate constant of palatal cells in untreated animals, and t' is the time to complete the ith stage of development. Note that two parameters, α and γ, must be estimated for growth rate of the cells in the palatal region, as well as an estimate of the background incidence of clefting in control litters. For cross species extrapolation, it would be necessary to further assume that postulated relationships between administrated dose and target dose and with cell kinetics and morphogenesis are similar in the exposed and target species. They cautioned that, even in this simplified case, many assumptions had to be made and the estimates of response from the BBDR model may not improve those obtained by logistic analysis of standard bioassay data. It is also worth noting that this approach did not consider among-animal variability in parameters and thus did not include an important component of modeling.

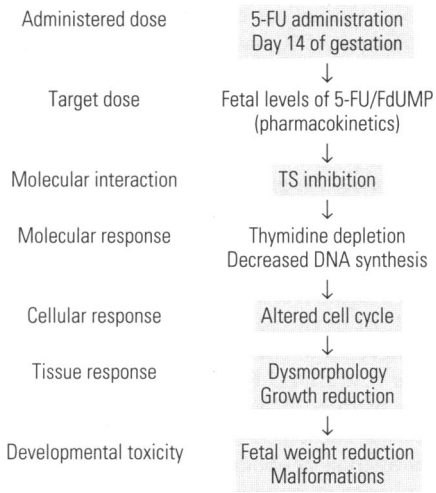

Figure 3. Framework of BBDR and application to 5-FU. Data from Shuey et al. (*19*).

In an attempt to prospectively build a BBDR model, Shuey et al. (*19*) examined the response of gestation day (GD) 14 rat fetuses to the chemotherapeutic 5-FU. This agent was selected because steps in altering cellular biochemistry are generally known (inhibition of the enzyme thymidylate synthetase by the metabolite 5-fluoro-2′-deoxyuridylic monophosphate) and the consequences of depleted nucleotides on the cell cycle, and hence fetal growth and hind limb development, were quantifiable. The experimental design included examination of a wide range of dose levels, tissues, end points, and time points following dosing. The end points included activity of thymidylate synthetase, the synthesis of DNA and protein, cell cycle kinetics, tissue morphometrics, and fetal morphology. In constructing the quantitative dose–response model, they focused on enzyme activity 1 hr after dosing, the percent of cells in S phase 8 hr after dosing, limb bud morphometry 24 hr after dosing, and the incidence of malformations on day 21 of gestation (7 days after dosing). Each step was described by a Hill equation; the individual equations were linked as proposed in the initial framework to yield an empirical model of the induction of developmental toxicity (Figures 4 and 5). The linked model slightly overestimated the incidence of hindlimb defects observed in the original bioassay. A Monte Carlo simulation to evaluate the amount of variability around the predicted relationship between the administered dose of 5-FU and the incidence of digit agenesis showed that a few combinations provided predicted curves near the original dose response. The model, though biologically based, must still be regarded as empirical because of the lack of *a priori* biological basis for the form of the quantitative expressions. As a consequence, the proposed model cannot be used to make predictions about other thymidylate synthetase inhibitors, effects following exposure at other critical times during development or via other routes of exposure, or in other species. Nevertheless, the effort provided clear evidence of the feasibility of constructing mechanistic models, as well as the somewhat daunting data intensiveness required of even relatively simple postulated cause-and-effect mechanisms. If experimental situations can be constructed in which portions of the response model can be verified with *in vitro* tissues from the test species and from human embryo tissues, then it might be possible to make the models less empirical in nature. In addition, if models could be built upon final common pathways of chemical perturbation (altered nucleic acid pools or altered cell cycle kinetics), then their use could be extended beyond the particular chemical under study. Leisenring et al. (*25*) have proposed such a model based upon cell kinetics and a branching model. In any event, the data intensiveness involved in the construction of models of this nature will, for the foreseeable future, limit their application to situations in which large segments of the population are exposed to low levels and where determination of safe exposure levels is extremely important or to economically important chemicals where the costs of regulation would warrant a concerted effort to minimize the uncertainties inherent in extrapolation. To generalize the predictiveness of the experimental model, a parallel modeling effort was undertaken using computer simulations of the toxicokinetics, cellular biochemistry, and cell kinetics.

Computer Simulation

The next step in the evolution of the dose–response model is to develop mathematical descriptions that capture the mechanisms that are responsible for the steps in the causal cascade from exposure to final effect. This has several advantages. If there is a good mapping between the biological structures and processes being modeled and the equations and parameters in the

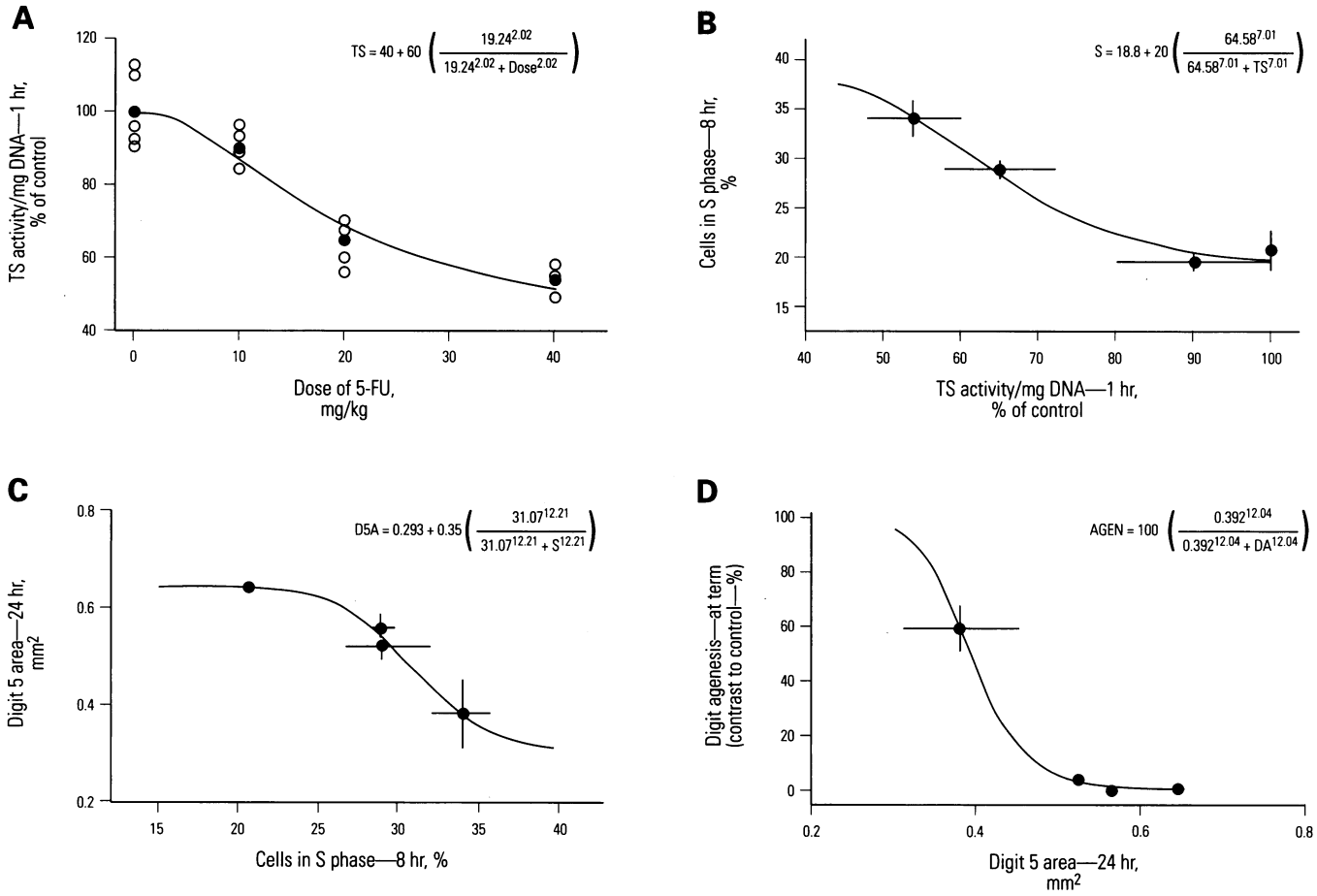

Figure 4. Relationships (described mathematically by Hill equations) between successive biochemical and cellular events in the fetal rat hindlimb following 5-FU exposure. TS, thymidylate synthetase. *A*, TS inhibition 1 hr after dosing; *B*, altered cell cycle following TS inhibition; *C*, growth reduction as a result of an altered cell cycle; *D*, malformations following growth reduction. Reproduced with permission from Shuey et al. (*19*).

model, and the biological processes in the animal for which the model is developed reflect the processes going on in the target species (usually humans), then interspecies extrapolation can be carried out by replacing the species-specific parameters. Under current dose–response methodology, there is little empirical support for low-dose extrapolation; however, the creation of the sort of biologically based dose–response model considered here presumes mechanistic understanding of the toxic processes, and lends credence to any low-dose extrapolation. A similar argument applies to route-to-route extrapolation. Finally, such a model incorporates data from a wider diversity of sources than conventional dose–response models. Dose–response bioassays and virtually any other relevant biological information could be incorporated into such models. In the development of such models, it is helpful if their structure reflects the biological structure of the system being modeled. Thus, it may help to divide the events to be modeled into steps as depicted in Figure 3, with each submodel linked together by common variables.

Conventional dose–response models relate some measure of final outcome, such as incidence of malformations or average weights, directly to an administered dosage. It is more natural for this kind of BBDR to predict time courses of phenomena because it is usually easier to express models in terms of rates. So, for example, a pharmacokinetic submodel will predict the concentration of a compound at the target cell surface for a continuous range of times after dosing; submodels for the interaction with molecular receptors and subsequent changes in cellular behavior will predict the time course of such behaviors (e.g., the fraction of cells undergoing apoptosis at a given time). Lastly, the time course of the final end points in the causal cascade must be related to the observed adverse effect.

Consider as an example a model for the effects of 5-FU on cell-cycle kinetics in the rat fetus described above (*19*). A mathematical model is being developed to facilitate the understanding of the relationships among enzyme activities, nucleotide concentrations, and subsequent perturbations of cell kinetics observed in this system. Although the model is still in development (*26–28*), it is far enough along to allow discussion of the salient points in nonmathematical form.

The mathematical model is an attempt to quantify an admittedly simple conceptual model for the developmental toxicity of 5-FU: when pregnant dams are dosed on GD14 by subcutaneous injection, the 5-FU is absorbed from the site of injection into a large, metabolically inactive compartment. From there it enters the blood and is distributed to a metabolically active compartment, which includes renal elimination and the uterus. From there, 5-FU

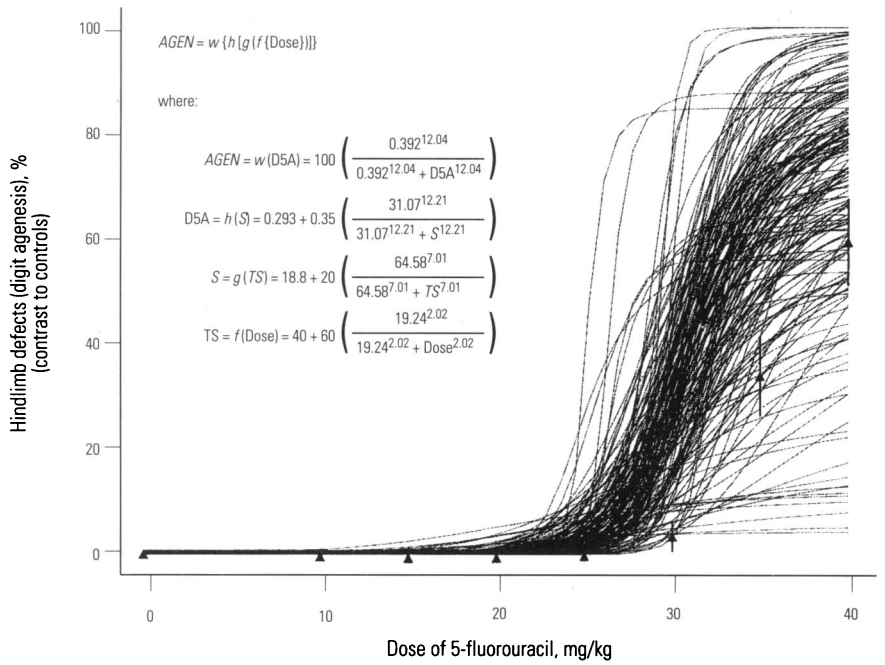

Figure 5. The integrated empirical model for 5-FU developmental toxicity based on TS inhibition. Individual curves were generated by Monte Carlo simulation to evaluate the amount of variability around the predicted relationship. Triangles (with lines) are the mean and standard errors of the original data. Reproduced with permission from Shuey et al. (*19*).

distributes uniformly throughout the uterine contents (fetuses), which are also metabolically active. Up to this point, the pharmacokinetic model is based on a model of Collins et al. (*29*) and modified to incorporate the uterine compartment after the model of O'Flaherty et al. (*30*). Parameters were estimated by fitting to data of Boike et al. (*31*), as well as empirically measured serum levels of 5-FU in the GD14 dam. A fraction of the metabolism of 5-FU results in FdUMP, the inhibitor of thymidylate synthetase (TS). The inhibition of TS and its consequences are modeled using a modified version of the equations published by Jackson (*32*) for the regulation of the composition of the deoxyribonucleoside triphosphate composition. The modification consisted of, first of all, rewriting the portion of the model dealing with TS in the form of mechanistic equations instead of the rapid equilibrium equations and allowing enzyme activities to change during the cell cycle [e.g., TS activity increases throughout S phase; (*33*)]. This latter point was deemed important because the biochemical kinetic system is highly nonlinear, and it was of interest to see to what extent embedding the system into an active cell cycle would alter its behavior. From this model we obtain predictions about the nucleotide pool sizes, rates of DNA synthesis and cell cylcle kinetics as a function of time, and concentration of FdUMP.

The 5-FU model as described includes only two of the kinds of submodels listed above: a pharmacokinetic model to predict 5-FU and FdUMP concentrations in fetal cells and a biochemical model that includes inhibition of TS (the receptor for FdUMP) by FdUMP and the resulting perturbations of DNA synthesis and cell cycle kinetics. Nevertheless, implementation of the model involves 17 differential equations and 65 parameters, virtually all of which were set to values estimated in the literature for similar mammalian systems. Only the fraction of 5-FU metabolized to FdUMP and the intracellular half-life of FdUMP were adjusted to match model predictions of the maximum inhibition of TS to our *in vivo* high-dose observations.

The original motivation for building this particular model was to account for an apparent discrepancy between the relationship of 5-FU dose to maximum TS inhibition on one hand and that of dose to the incidence of malformations and weight deficits on the other (*19*). In brief, as the dosage increases, the marginal increase of TS inhibition decreases sharply. The dose response for malformations has nearly a hockey-stick shape (Figure 2*A*); at the dosages at which malformations begin to increase, the additional increase of TS inhibition due to increases in 5-FU dose has already become very small. The mathematical model reproduces this behavior (Figure 6). Thus, what at first seemed to be evidence for the action of mechanisms other than initially hypothesized is actually the normal behavior of a highly nonlinear interactive system.

The 5-FU model as described here is far from being useful as a quantitative dose–response model for risk assessment; however, it serves as the source of some instructive points. First of all, the model reproduces the very steep increase in malformations seen in rats exposed *in vivo* beginning at 30 mg/kg in its prediction of cell cycle disruption (Figure 7). Although this may yet turn out to be an artifact of model construction, the observation points out the possibility of using such models to infer safe exposure levels based on a quantitative understanding of the mechanism of action of a toxicant. By looking at how much the threshold varies when the model is run with different values for the kinetic parameters, it would be possible to get at least a semiquantitative estimate of the degree of intraspecies variability in the threshold one might expect. Thus the model could serve as a link between the variability of metabolic values, which could be observable experimentally, and the uncertainty factor for intraspecies variability, which so far is usually determined purely formally. Finally, by putting hypotheses about mode of action in a quantitative language, experimental tests of those hypotheses can be carried out in a more rigorous manner, facilitating decisions about the appropriateness of an animal model for

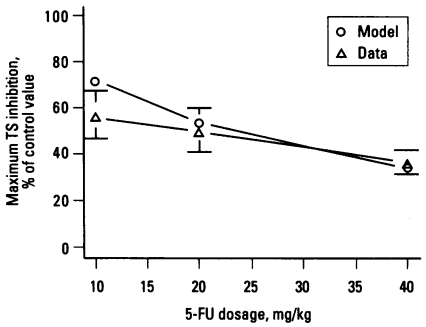

Figure 6. Maximum TS inhibition in whole rat embryos dosed on gestation day 14 compared to predictions from the model. Values shown are expressed as percent of control and 95% confidence intervals.

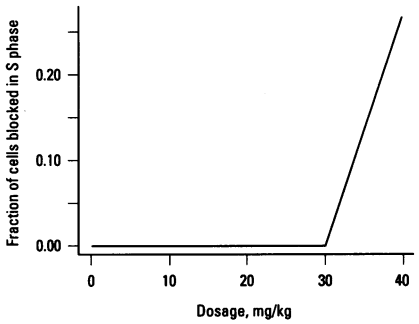

Figure 7. Model-predicted dose response for the fraction of cells blocked in the cell cycle.

extrapolation to humans. In addition to refining the precise mathematical form of the model, we are currently evaluating several model-generated hypotheses concerning the effects on nucleotide pool sizes, cell cycle kinetics, and rescue by presumably rate-limiting nucleotides in a simpler *in vitro* cell system. If these experiments are successful, similar information will be collected on embryos exposed *in utero*. All along, there will be an iterative interplay between model formulation and experimental data as we learn more about the underlying biological processes.

It is reasonable to inquire whether the effort implied by this discussion of BBDR modeling could ever be justified in routine practice. Clearly, in our current state of knowledge, it would be unrealistic in the extreme to propose that all dose–response assessments for developmental toxicity should be conducted under this paradigm. However, there may be toxic agents with such pervasive distribution and potentially toxic effects that such intense effort is justified: chemicals such as dioxins, PCBs, and other persistent environmental chemicals that mimic or inhibit the effect of endogenous hormones may provide examples. Such efforts may also be warranted for chemicals that are being proposed for significantly new uses in which the potential economic gain to the supplier and resulting widespread human exposure justifies extended examination of the hazards identified by more traditional toxicological approaches (e.g., fuel additives, alternative fuels). Perhaps the greatest potential for such models is not for assessments of specific chemicals but as research tools to help elucidate general mechanisms of toxicity. As the experience with such investigations accumulates, it may be that their results could be treated as a toolbox out of which models for new compounds could be constructed with only incremental increases in effort. In the meantime, the rigor required to construct BBDR models can only benefit toxicological mechanistic investigation.

The Embryologically Based Approach

The final approach we present is clearly the most visionary and hypothetical of the approaches. The EBDR approach begins not with an effect and mechanism but with the fundamental understanding of normal morphogenesis and, only secondarily, factors in how these events are perturbed by exogenous agents. Such models would be adaptable to the effects of multiple chemicals provided they captured the salient biological events. The biological understanding of morphogenesis at the molecular level, linked with mathematical theories and constructs of pattern formation, may open the door for these approaches. We will use recent advances in the understanding of the role of homeobox genes in development of the axial skeleton and limbs to illustrate how the emerging knowledge of positional signaling is providing information to take the heretofore theoretical models of pattern formation into potential use by developmental toxicologists.

Homeobox genes, so named for a 183 bp DNA sequence that yields a 61 amino acid protein sequence containing a DNA binding domain, were first described in *Drosophila* about 10 years ago (*34*). These genes are highly conserved across many phyla, and today some 38 *Hox* genes organized in four chromosome complexes are recognized in the mouse (Figure 8). The role of *Hox* genes in pattern formation in vertebrates is in part due to a feature termed collinearity, that is, there is a close relationship between expression along the anterior–posterior axis of the embryo and the gene's order along the chromosome (Figures 9 and 10) (*35–37*). Manak and Scott (*36*) provided several conserved rules governing *Hox* gene function in the developing vertebrae: *a*) their tissue expression and function follows along the order on the chromosome; *b*) more genes are expressed in the more posterior regions; *c*) loss of gene function leads to development of anterior structures where posterior structures should have formed (e.g., a rib on the 14th postcervical vertebra of a rodent); *d*) activation of genes where they should be off (gain-of-function mutations) leads to formation of posterior structures where anterior structures should be (e.g., the presence of only 12 pairs of ribs in rodents); *e*) each homeotic gene contains a single homeobox, which encodes a specific DNA-binding transcription factor; and *f*) most of the 5′ ends of the transcription are oriented toward the end of the cluster. The upstream and downstream events from Hox expression have yet to be generally established (*38,39*), but their

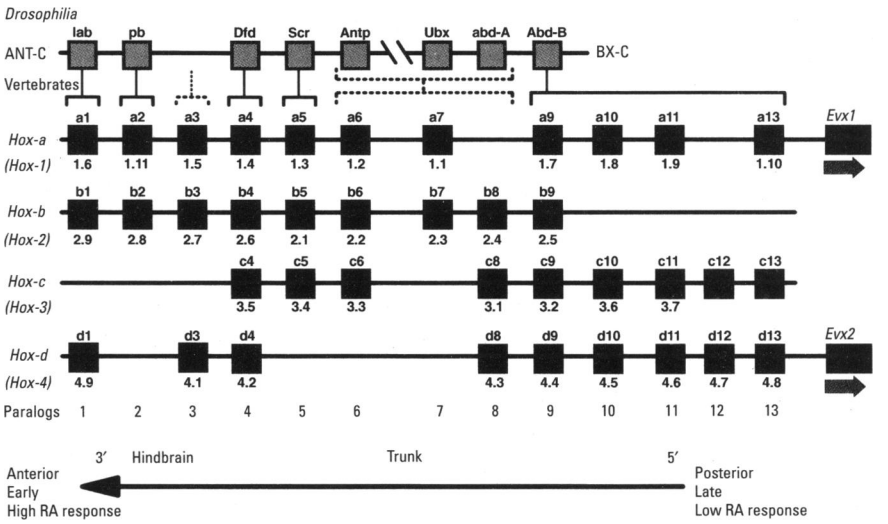

Figure 8. Alignment of the vertebrate Hox clusters into paralogous groups and comparison with *Drosophila HOM-C*. The 13 paralog groups are labeled at the bottom. In each of the four vertebrate clusters, the boxes denote known homeobox genes; above each box is the current nomenclature. The older nomenclature is denoted beneath each box. The stippled columns denote homology relationships that exist with specific *Drosophila* genes. The genes at the left are expressed earlier in development and have the highest sensitivity to retinoic acid (RA). Modified from Krumlauf (*35*).

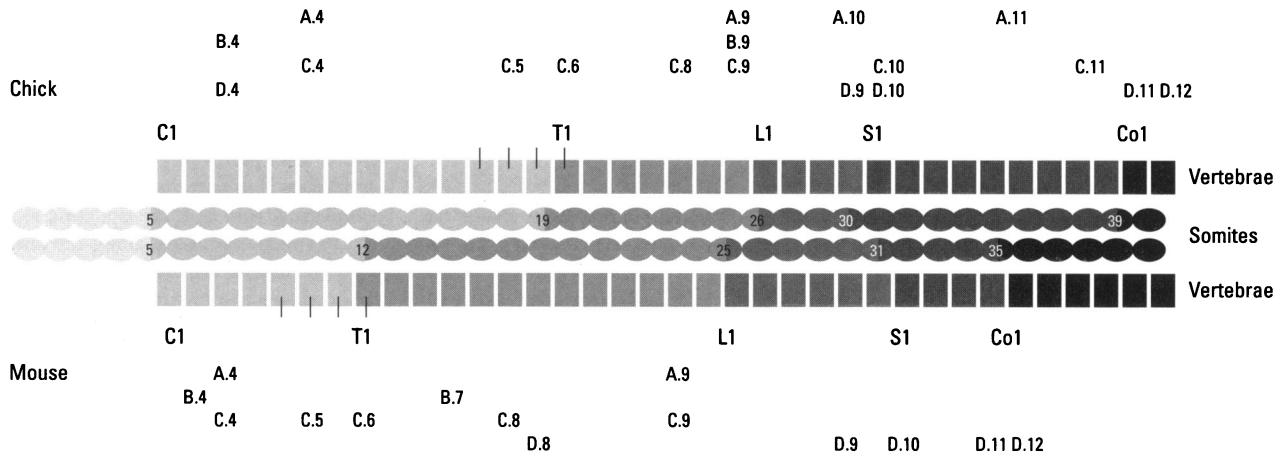

Figure 9. Schematic representation of the axial formulae, somite levels, and anterior expression boundaries of *Hox* genes in the paraxial mesoderm of the chick and mouse. Note that the 3′ paralogs (lower numbers) have more anterior boundaries than those from the 5′ end. Modified from Burke et al. (*37*)

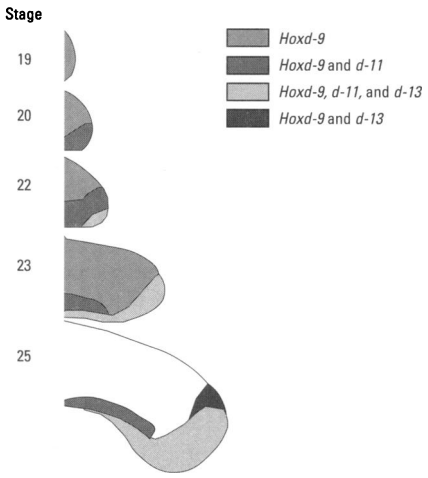

Figure 10. Schematic representation of the *Hox* gene expression in the chick leg bud at various Hamburger and Hamilton stages. The view is of the dorsal surface of the right limb bud. *Hox-10* and *Hox-12* have been omitted for the sake of clarity. Note the progressive expression of the 5′ genes as development proceeds. Modified from Morgan and Tabin (*73*).

coordinated appearance during early pattern formation implicates them in at least specifying segment identity if not the actual segmentation process. Other gene classes are also known to be involved in sculpting the formation of the vertebrae, with the *Pax* genes (especially those that include the paired box and homeobox) among the more well studied (*40*).

During the last several years, a number of loss-of-function mutants created by recombinant genetic techniques illustrate the governing rules stated above (Table 1). Phenotypes similar to those of the knockout experiments are familiar to developmental toxicologists, and it should not be surprising to see that homeotic-type alterations in the vertebrae, particularly the anterior and posterior boundaries of ribs, are frequently induced by xenobiotics. Thus, anteriorizations (i.e., the taking on of the morphology of the immediately preceeding segment) of the first lumbar vertebrae as induced by maternal toxicity or stress (*58,59*); valproic acid (*60*); bromoxynil (*61,62*); salicylate (*63*); dimethadione (*64*) and retinoic acid (*65*) are but a few examples. Posteriorizations of the thoraco–lumbar border have also been observed (*66–68*), but this effect is clearly less frequent in the developmental toxicology literature. Toxicant-induced alterations are not limited to the thoraco–lumbar border and may also involve posteriorizations of the cervical–thoraco juncture, as demonstrated by experiments with methanol (*69,70*) and boric acid (*67*). In general, these agents induce frank dysmorphologies of the axial skeleton at higher dose levels; however, the phenotypes rarely, if ever, completely resemble the null hox phenotypes regardless of response level. Closer examination of subtle skeletal morphological features in developmental toxicity bioassays may strengthen the impact of the effects on the axial skeleton. We see changes at the major boundaries because that is where our attention has been focused and because of the ease of observing such changes. In addition, comparison with phenotypes from individual null mutations might not be expected to produce complete concordance due to other possible pathogenic pathways as well as the potential redundancy in *Hox* gene function (*52*).

The study of pattern formation in the limb has perhaps received the most attention of any vertebrate organ (*71–75*). The limb begins as an outgrowth of the lateral somatic mesoderm as cell proliferation slows in the regions immediately anterior and posterior to the emerging bud. An apical epidermal ridge (AER) soon develops and is maintained by the underlying mesoderm. In turn, the AER supports proliferation of the underlying mesenchymal cells in what is termed the progress zone (PZ). Recent evidence suggests that fibroblast growth factor (FGF) 2 and FGF 4 may be the morphogenetic signal emanating from the AER. Continued cell proliferation in the PZ gradually establishes the proximal–distal axis of the limb. As shown by surgical removal of the AER at various stages, the longer cells stay in the PZ, the more distal the structures they will form. As cells leave the PZ, they decrease their rate of proliferation and begin to differentiate; thus as they exit, they are believed to have acquired some aspects of positional identity. A small set of cells in the posterior area of the PZ then gains a special property that contributes to the anterior–posterior axis. This area, known as the zone of polarizing activity (ZPA), was first identified by its ability to induce digit duplications when grafted into the anterior region of the emerging bud (*76*). The polarizing ability of the ZPA was at one time thought to be due to its role in establishing a gradient of retinoic acid, but it has now been demonstrated to result from expression of *sonic hedgehog* (*Shh*), a gene related to *Drosophila* segment polarity gene *hedgehog* (*77*). The *Shh* gene product is an autocleavable protein whose active amino terminus

Table 1. Phenotypes of null mutations for paralogs in the four vertebrate *Hox* clusters, with particular reference to specification of the axial skeleton.

Paralog	*Hoxa*	*Hoxb*	*Hoxc*	*Hoxd*
1	No rhombomeres; ear defect; poor form of ganglions IX and X, posteriorization of ganglions VII and VIII (*41*)		NA	
2	Cleft of secondary palate; defects in inner, middle, and outer ear; cleft in root of tongue; reduced size of facial nerve (*42,43*)		NA	NA
3	Athymic; aparathyroid; decreased thyroid and sub-maxillary glands; defects in heart, arterial, and cardio-pulmonary failure; lethal (*44,45*)		NA	Loss of dens axis; ectopic ossification of caudal edge of basioccipital bone; displacement of lateral foramina on C2 (*44,46*)
4	Non-normal dorsal process on C2 and C3; transformations of C3 to C2, and C7 to T1 (*47,48*)	Transformation of C2 to C1; sternal defects (*49*)		Transformation of C2 to C1; malformations of C1 to C3 (*50*)
5	C7 rib; no TA on C6; abnormal number of sternebrae and ribs (*51*)	Forelimbs shifted anteriorly relative to axial skeleton by 1 to 2 cervical vertebrae; anteriorization C6 thru T1 (*52*)		NA
6	Split in ventral T1; C7 rib; transformation of T1 to T2 (*48*)	Missing 1st rib; bifurcation of 2nd rib at the sternum; anteriorization C6 thru T1 (*52*)		NA
7			NA	NA
8	NA		8th pair of ribs on, instead of below, sternum; ribs on L1; die within days of birth (*53*)	
9			Anterior transformation of T10 through L1; 8 or 9 pairs of attached ribs (*54*)	
10		NA		
11	12 rib pairs; complete fusion of pisiform to triangular carpal bones (*55*)	NA		Anteriorization of sacral bones; fusions between wrist bones; defects in the radius and ulna; extra lumbar vertebrae (*56*)
12	NA	NA		
13		NA		Splits, fusions in digits, and extra digits (*57*)

NA, not applicable (paralog not discovered).

remains within the cell of origin and therefore does not meet the criterion of a true morphogen. Ectopically placed retinoic acid-soaked beads have been shown to induce *Shh*, create a new ZPA, and cause mirror-image digit duplications.

As with the axial skeleton, *Hox* genes are important in providing positional information that defines axes of the limb, if not of the digits themselves. Genes in the *Hoxa* cluster have been reported to emerge in a collinear fashion from the PZ, such that the more 3′ *Hoxa-11* is expressed more distally than *Hoxa-13* (*78*). This gradient may be involved in proximal–distal axis definition. Similarly, the *Hoxd* cluster is expressed in a collinear fashion in response to the ZPA, with *Hoxd-9* coming to be expressed in the most anterior and proximal region and *Hoxd-10, d-11, d-12,* and *d-13* having progressively more posterior and distal expressions. Expression of the *Hoxd* cluster in the limb is dependent on the presence of the AER. The progressive expression also correlates with the onset of asymmetry within the limb, as the posterior half becomes larger than the anterior as *Hoxd-11* and *Hoxd-12* are activated, perhaps suggesting a role in mediating growth (*73*). By providing this overlapping gradient of expression, the five members of the *Hoxd* cluster help define the number and placement of digit types. The fact that the *Hoxa* gradient is orthogonal to the *Hoxd* gradient may give further positional identity to cells. However, results of knockout experiments have not necessarily supported some facets of this model of positional information. For example, following disruption of *Hoxd-13* (which might be hypothesized to interfere with the most posterior digit based upon its expression pattern), all digits, but particularly digits II and V, are reduced in size, and half the animals possessed an additional rudimentary digit posterior to digit V (*57*). The early Hox gradients are also clearly dynamic in space and time, and the more restricted domains seem to fade with time into a more uniform expression pattern (*73*). As with the axial skeleton, the downstream events from Hox expression in the branching process and formation of bones in the limb remain to be discerned, although retinoic acid and its metabolites, binding proteins and receptors, are undoubtably involved (*72,79*). Finally, the *Wnt* genes appear to have a role in establishing the dorsal–ventral axis of the limb bud (*80*).

To what extent does our increasing knowledge of the molecular foundation of pattern formation allow us to judge the significance of toxicant-induced homeotic shifts or other structural perturbations during development? Are the *Hox* genes directly involved in providing positional

information for subsequent pattern formation? Can we identify alterations in patterning-gene expression in the immediate stages following toxicant exposure when the agent is initiating the morphologic lesion? The difficulty in establishing relative landmarks at these early embryonic stages and quantitating molecular responses at the cellular level makes this problematic. Are the shifts in the boundaries the limit of their phenotypic expressions or just the tip of other responses that are more difficult to identify? If direct primary links between xenobiotics and altered gene function become evident, will we become more worried because we are perturbing the action of transcription factors key to morphogenesis; or do we become less worried as we understand the redundancies built into the overlapping expression of paralog members? Before we can answer such questions, we need to be able to quantify expression in space and time on the cellular level, characterize the extent of variability in controls, learn exactly how xenobiotics can affect expression (do they all, for instance, modulate local retinoic acid concentrations?), and understand the downstream events that translate the expression into cellular characteristics of particular segments (be it vertebrae or limb components). In these regards, it is desirable to have frameworks in place for assimilating the information and even testing hypotheses using computer models, as is being done for the 5-FU BBDR presented above. It is in this context that we introduce more theoretical formulations of pattern formation in biological systems.

Since Turing introduced the concept of morphogen in 1952 (*81*), modelers have explored the consequences of different theories of pattern formation through the behavior of mathematical models. Two major kinds of models have been devised to explain the complex patterns seen in development. The first type of model supposes that the development of pattern occurs in two stages. In the first stage, a field that provides positional information is laid down; it is a prepattern to be used by developing cells to determine their position in the field. During the second stage, cells sense their positions relative to this chemical coordinate system and react appropriately. This concept has been referred to as positional information (*82,83*). In the second type of model, the final pattern unfolds due to the manifestation of mechanical and chemical interactions among the developing cells. The volume by Othmer et al. (*84*) and the references cited therein provide a good introduction to this topic.

Models for developmental processes have been mostly abstract explanations of the consequences of hypothesized mechanisms of pattern formation. Molecular embryology, as exemplified by the discussion of the role of *Hox* genes and other factors in specifying segment identity, is now providing concrete expressions for the hypothesized mechanisms. Thus, models of pattern formation should take on new significance when combined with the detailed understanding of how locational information is actually expressed at the molecular level during development.

Historically, two principal mechanisms have been proposed to convey positional information in the mathematical models. The most common approach uses a gradient of chemicals to provide the positional information. In an interesting recent example, Levin (*85*) supposes that the products of two genes are initially distributed on a gradient and interact intracellularly to produce complicated patterns, making use of fractals and chaos theory to generate developmental patterns. A more common model relies on pairs of diffusing chemicals that interact with each other to form stable patterns in the concentration of the two chemicals (reaction–diffusion models). In the general form for such a system, one of the chemicals catalyzes its own synthesis as well as that of a second chemical, which inhibits the synthesis of the first chemical. Both chemicals diffuse away from their point of synthesis, but the inhibitor diffuses faster. Thus, there arises a common developmental pattern of local self-enhancement and long-range inhibition. Simple models, with this structure as a point of departure and with a single spatial dimension as well as time, can demonstrate a number of spatial patterns such as gradients or stripes and can qualitatively reproduce the behavior of some experimental developing systems after they are perturbed (*86*).

In the other approach, mechanico–chemical models, pattern formation and morphogenesis evolve simultaneously to produce the final pattern. Here a more simple positional signal, for example, a simple gradient in a chemical, interacts with other mechanical aspects of cells, such as differential adhesion and motility, to generate the final pattern. No prepattern is formed in these models; rather, it is the innate behavior of the cells themselves that forms the final morphogenetic pattern. As in reaction–diffusion models, successful mechanico–chemical models generate a pattern of local enhancement and long-range inhibition (*87*). Figure 11 shows an example of how a simple model involving chemotaxis and mitosis of pigment cells can reproduce the complicated skin pigmentation patterns seen in snakes (*88*). Sometimes, to account for the development of complex patterns, the underlying parameters of the model, such as the size of the developing tissue or organ, are allowed to change (*87,88*). The resulting pattern is an interaction between the stable pattern that would have evolved for fixed parameters and the change in the parameters.

For much mathematical modeling in developmental biology, the general form has been to show that some specific pattern could be generated by a particular mechanism. Often, the absence of much-detailed biological information, such as rates of reaction, actual cellular behavior, and even the identity of hypothetical reacting chemicals, has forced such models to be fairly abstract. Even so, their ability to show the consequences of simple hypotheses of interaction has been valuable, both for testing hypotheses about mechanisms of development and for augmenting biologists' intuition about such systems. However, in risk assessment, we are likely to have fairly detailed and specific information about how a chemical interacts at the subcellular level and how the cells' behavior changes as a consequence. To be able to link these changes to developmental changes, more detailed and specific models for morphogenesis are needed. The problem changes from demonstrating that a particular pattern could be generated by a given mechanism to quantifying how much a particular cellular behavior can change without affecting subsequent morphogenesis or, more generally, how much morphogenesis is affected by a given change in cellular behavior. Naturally, before we can solve this problem, much more needs to be learned about normal development, and more specific models for normal development need to be developed. This is perhaps an area where developmental toxicologists and developmental biologists can collaborate successfully. By using the plethora of developmental perturbations available through developmental toxicants and observing and modeling their effects on normal development, our knowledge of both normal and abnormal development should be greatly enhanced. The emerging knowledge base on molecular morphogenesis as exemplified by the axial skeleton and limbs appears ripe for the task.

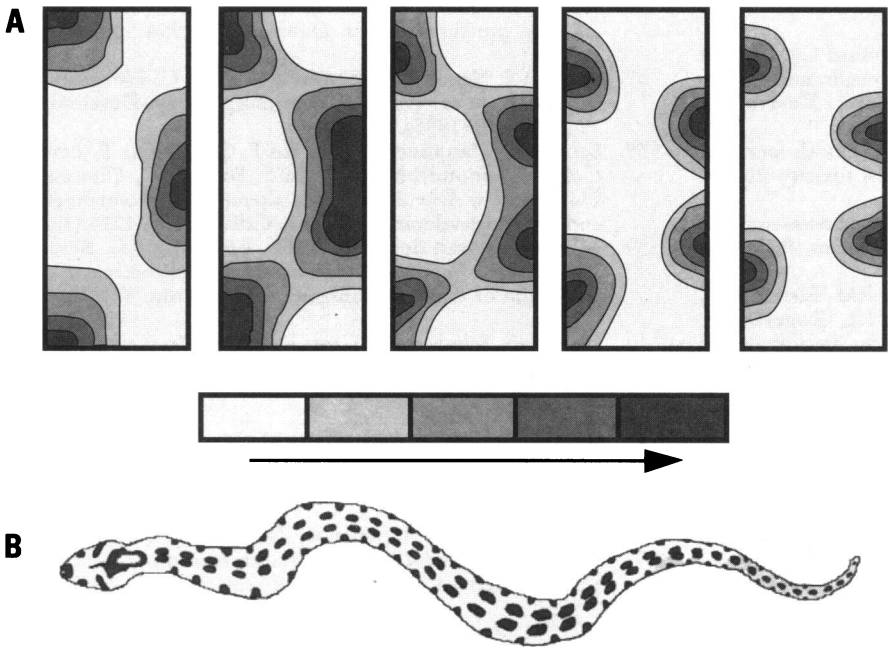

Figure 11. Patterns produced by the model mechanism of chemotaxis and mitosis of pigment cells can reproduce the complicated pigment patterns in snakes. The following equations describe this mechanism:

rate of change of cell density = diffusion − chemotaxis + cell mitosis

$$\partial n/\partial t = D\nabla^2 n - \alpha\nabla\bullet(n\nabla c) = sm(N-n)$$

rate of change of chemoattractant = diffusion + chemoattractant production − degradation

$$\partial c/\partial t = \nabla^2 c + s[n/(1+n) - c].$$

A, shows how the pattern changes as the chemotactic parameter α increases in the range 19.92 to 63.43. *B*, shows *Elaphe situla* with paired spots. Modified from Murray and Myerscough (*88*).

Conclusions

Risk assessment for developmental end points is entering a state of flux after a relatively long period of inertness. The benchmark dose concept is certainly an improvement over the use of NOAELs and LOAELs (lowest observed adverse effect levels) for standard setting. Nevertheless, it presents the same problems for extrapolation as the NOAEL. Only by developing more complete pictures of how developmental toxicants perturb normal development will we be able to put extrapolation on a more empirical and scientific footing. This will not be easy and it will require that developmental biologists become familiar with developmental toxicology (and vice versa), that biologists become more comfortable with quantitative methods, and that biomathematicians be willing to work with the often complex and inelegant mathematical systems required to more realistically model specific biological systems. We also need to develop better ways to use partially developed models to enhance more statistical approaches to risk assessment, perhaps by modifying uncertainty factors or through dose scaling, as we occasionally use pharmacokinetic models now.

REFERENCES

1. U.S. Environmental Protection Agency. Guideline for developmental toxicity risk assessment. Federal Register 56:63798–63826 (1991).
2. National Research Council. Risk Assessment in the Federal Government: Managing the Process. Washington:National Academy Press, 1983.
3. Barnes D, Dourson ML. Reference dose (RfD): description and use in health risk assessment. Regul Toxicol Pharmacol 8:471–486 (1988).
4. Moore JA, Daston GP, Faustman E, Golub MS, Hart WL, Hughes C, Kimmel CA , Lamb JC IV, Schwetz BA, Scialli AR. An evaluative process for assessing human reproductive and developmental toxicity of agents. Reprod Toxicol 9:61–95 (1995).
5. Clark RL, Robertson RT, Peter CP, Bland JA, Nolan TE, Oppenheimer L, Bokelman DL. Association between adverse maternal and embryo-fetal effects in norfloxacin-treated and food-deprived rabbits. Fundam Appl Toxicol 7:272–286 (1986).
6. Clark RL, Robertson RT, Minsker DH, Cohen S, Tocco J, Allen HL, Jones ML, Bokelman DL. Diflunisal-induced maternal anemia as a cause of teratogenicity in animals. Teratology 30:319–332 (1984).
7. Moore JA, IEHR Expert Scientific Committee. An assessment of lithium using the IEHR evaluative process for assessing the human developmental and reproductive toxicity of agents. Reprod Toxicol 9:175–210 (1995).
8. Gaylor D. Incidence of developmental effects at the no observed adverse effect level (NOAEL). Regul Toxicol Pharmacol 15:151–160 (1992).
9. Leisenring W, Ryan L. Statistical properties of the NOAEL. Regul Toxicol Pharmacol 15:161–171 (1992).
10. Crump K. A new method for determining allowable daily intakes. Fundam Appl Toxicol 4:854–871 (1984).
11. Barnes DR, Daston GP, Evans JS, Jarabek AM, Kavlock RJ, Kimmmel CA, Park C, Spitzer H. Benchmark dose workshop: criteria for use of a benchmark dose to estimate a reference dose. Regul Toxicol Pharmacol 21:296–306 (1995).
12. Faustman EM, Allen BC, Kavlock RJ, Kimmel CA. Dose–response assessment for developmental toxicity. I. Characterization of database and determination of the no observed adverse effect levels. Fundam Appl Toxicol 23:478–486 (1994).
13. Allen BC, Kavlock RJ, Kimmel CA, Faustman EM. Dose–response assessment for developmental toxicity. II. Comparison of generic benchmark dose estimates with no observed adverse effect levels. Fundam Appl Toxicol 23:487–495 (1994).
14. Allen B, Kavlock RJ, Kimmel CA, Faustman EM. Dose–response assessment for developmental toxicity. III. Statistical models. Fundam Appl Toxicol 23:496–509 (1994).
15. Kavlock RJ, Allen BC, Faustman EM, Kimmel CA. Dose–response assessment for developmental toxicity. IV.

Benchmark doses for fetal weight changes. Fundam Appl Toxicol 26:211–222 (1995).
16. Catalano PJ, Scharfstein DO, Ryan LM, Kimmel CA, Kimmel GL. Statistical model for fetal death, fetal weight, and malformation in developmental toxicity studies. Teratology 47:281–290 (1993).
17. Zhu Y, Krewski D, Ross WH. Dose–response models for correlated multinomial data from developmental toxicity studies. Appl Stat 43:583–598 (1994).
18. Krewski D, Zhu Y. Applications of multinomial dose–response models in developmental toxicity risk assessment. Risk Anal 14:595–609 (1994).
19. Shuey DL, Lau C, Logsdon TR, Zucker RM, Elstein KH, Narotsky MG, Setzer RW, Kavlock RJ, Rogers JM. Biologically based dose–response models for developmental toxicology: biochemical and cellular sequelae of 5-fluorouracil exposure in the developing rat. Toxicol Appl Pharmacol 126:129–144 (1994).
20. Kavlock RJ, Schmid J, Setzer RW. A simulation study of the influence of study design on the estimation of benchmark doses for developmental toxicity. Risk Anal (in press).
21. Tukey J, Ciminera J, Heyse B. Testing the statistical certainty of a response to increasing dose of a drug. Biometrics 41:295–301 (1985).
22. Conolly RB, Andersen ME. Biologically based pharmacodynamic models. Annu Rev Pharmacol Toxicol 31:503–523 (1991).
23. Frederick C. Limiting the uncertainty in risk assessment by development of physiologically based pharmacokinetic models. Toxicology 68:159–175 (1993).
24. Gaylor D, Razzaghi M. Process of building biologically based dose–response models for developmental defects. Teratology 46:573–581 (1992).
25. Leisenring WM, Leroux BG, Moolgavkar SH, Faustman EM. Evaluation of cellular kinetics in biologically based dose–response models for developmental toxicity. Teratology 47: 426 (1993).
26. Setzer RW. Development of biologically based dose–response models: modeling the effects of 5-FU on cell cycle kinetics. Teratology 47:435–436 (1993).
27. Setzer RW. Development of biologically based dose response models: using a mechanistic model to generate experimentally testable hypotheses about mechanisms of developmental toxicity of 5-FU. Teratology 49:396–397 (1994).
28. Lau C, Copeland MF, Mole L, Setzer RW, Rogers JM, Kavlock RJ. Development of a biologically based dose–response (BBDR) model for the embryotoxicity of 5-fluorouracil (5-FU) in the rat. Teratology 51:188 (1995).
29. Collins JM, Dedrick RL, King FG, Speyer JL, Myers CE. Nonlinear pharmacokinetic models for 5-fluorouracil in man: intravenous and intraperitoneal routes. Clin Pharmacol Ther 28:235–246 (1980).
30. O'Flaherty EJ, Scott W, Schreiner C, Beliles RP. A physiologically based kinetic model of rat and mouse gestation: disposition of a weak acid. Toxicol Appl Pharmacol 112:245–256 (1992).
31. Boike GM, Deppe G, Young JD, Malone JM, Malviya VK, Sokol RJ. Chemotherapy in a pregnant rat model. 2. 5-Fluorouracil: nonlinear kinetics and placental transfer. Gynecol Oncol 34:191–194 (1989).
32. Jackson RC. A kinetic model of regulation of the deoxyribonucleoside triphosphate pool composition. Pharmacol Ther 24:279–301 (1984).
33. Nagarajan M, Johnson LF. Regulation of thymidylate gene expression in mouse fibroblasts synchronized by mitotic selection. Exp Cell Res 181:289–297 (1989).
34. McGinnis W, Garbar RL, Wirz J, Kuroiwa A, Gehring WJ. A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. Cell 37:403–408 (1984).
35. Krumlauf R. *Hox* genes in vertebrate development. Cell 78:191–201 (1994).
36. Manak JR, Scott MP. A class act: conservatism of homeodomain protein functions. Development 1994 Suppl: 61–71 (1994).
37. Burke AC, Nelson CE, Morgan BA, Tabin C. *Hox* genes and the evolution of vertebrate axial morphology. Development 121:333–346 (1995).
38. Schneider-Maunoury S, Topilko P, Setianidou T, Levi G, Cohen-Tannoudji M, Pournin S, Babinet C, Charnay P. Disruption of *Krox-20* results in alteration of rhombomeres 3 and 5 in the developing hindbrain. Cell 75:1199–1214 (1993).
39. Alkema MJ, van der Lugt NMT, Bobeldijk RC, Berns A, van Lohuizen M. Transformation of axial skeleton due to overexpression of *bmi*-1 in transgenic mice. Nature 374:724–727 (1995).
40. Dietrich S, Schubert FR, Gruss P. Altered *Pax* gene expression in murine notochord mutants: the notochord is required to initiate and maintain ventral identity in the somite. Mech Dev 44:189–207 (1993).
41. Lufkin T, Dierich A, LeMeur M, Mark M, Chambon P. Disruption of the *Hox-1.6* homeobox gene results in defects in a region corresponding to its rostral domain of expression. Cell 66:1105–1119 (1991).
42. Gendron-Maguire M, Mallo M, Zhang M, Gridley T. *Hoxa-2* mutant mice exhibit homeotic transformation of the skeletal elements derived from the neural crest. Cell 75:1317–1331 (1993).
43. Rijli FM, Mark M, Lakkaraju S, Dierich A, Dolle P, Chambon P. A homeotic transformation is generated in the rostral branchial region of the head by disruption of *Hoxa-2*, which acts as a selector gene. Cell 75:1333–1349 (1993).
44. Condie BG, Capecchi MR. Mice with targeted disruptions in the paralogous genes *hoxa-3* and *hoxd-3* reveal synergistic interactions. Nature 370:304–307 (1994)
45. Chiska O, Capecchi MR. Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *hox-1.5*. Nature 350:473–476 (1991).
46. Condie BG Capecchi, MR. Mice homozygous for a targeted disruption of *Hoxd-3* (*Hox-4.1*) exhibit anterior transformations of the first and second cervical vertebrae, the atlas and axis. Development 119:579–595 (1993).
47. Horan GSB, Wu K, Wolgemuth DJ, Behringer RR. Homeotic transformation of the cervical vertebrae in *Hoxa-4* mutant mice. Proc Natl Acad Sci 91:12644–12648 (1994).
48. Kostic D, Capecchi MR. Targeted disruptions of the murine *Hoxa-4* and *Hoxa-6* genes result in homeotic transformations of components of the vertebral column. Mech of Dev 46:231–247 (1994).
49. Ramirez-Solis R, Zheng H, Whiting J, Krumlauf R, Bradley A. *Hoxb-4* (*Hox-2.6*) mutant mice show homeotic transformation of a cervical vertebra and defects in closure of the sternal rudiments. Cell 73:279–294 (1993).
50. Horan GSB, Nagy Kovacs E, Behringer RR, Featherstone MS. Mutations in paralogous *Hox* genes result in overlapping homeotic transformations of the axial skeleton: evidence for unique and redundant function. Dev Biol 169:359–372 (1995).
51. Jeannotte L, Lemieux M, Charron J, Poirier F, Robertson, EF. Specification of axial identity in the mouse: role of the *Hoxa-5* (*Hox1.3*) gene. Genes Dev 7:2085–2096 (1993).
52. Rancourt DE, Tsuzuki T, Capecchi MR. Genetic interaction between *hoxb*-5 and *hoxb*-6 is revealed by nonallelic noncomplementation. Genes Dev 9:108–122 (1995).
53. Le Mouellic H, Lallemand Y, Brulet P. Homeosis in the mouse induced by a null mutation in the *Hox-3.1* gene. Cell 69:251–264 (1992).
54. Suemori H, Takahashi N, Noguchi S. *Hoxc-9* mutant mice show anterior transformation of the vertebrae and malformation of the sternum and ribs. Mech Dev 51:265–273 (1995).
55. Small KM, Potter SS. Homeotic transformations and limb defects in *Hoxa-11* mutant mice. Genes Dev 7:2318–2328 (1993).
56. Davis AP, Capecchi MR. Axial homeosis and appendicular skeleton defects in mice with a targeted disruption of *hoxd-11*. Development 120:2187–2198 (1994).

57. Dolle P, Dierich A, LeMeur M, Schimmang T, Schuhbar B, Chambon P, Duboule D. Disruption of the *Hoxd-13* gene induces localized heterochrony leading to mice with neotonic limbs. Cell 75:431–441 (1993).
58. Kavlock RJ, Chernoff N, Rogers E. The effect of acute maternal toxicity on fetal development in the mouse. Teratog Carcinog Mutagen 5:3–14 (1985).
59. Chernoff N, Miller DB, Rosen MB, Mattscheck CL. Developmental effects of maternal stress in the CD-1 mouse induced by restraint stress on single days during the period of major organogenesis. Toxicology 51:57–65 (1988).
60. Narotsky MG, Francis EZ, Kavlock RJ. Developmental toxicity and structure activity relationships of aliphatic acids, including dose-response assessment of valproic acid in mice and rats. Fundam Appl Toxicol 22:251–265 (1994).
61. Rogers JM, Francis BM, Barbee BD, Chernoff N. Developmental toxicity of bromoxynil in mice and rats. Fundam Appl Toxicol 17:442–447 (1991).
62. Chernoff N, Rogers JM, Turner CI, Francis BM. Significance of supernumerary ribs in rodent developmental toxicity studies: postnatal persistence in rats and mice. Fundam Appl Toxicol 17:448–453 (1991).
63. Kimmel CA, Wilson JG. Skeletal deviation in rats: malformations or variations? Teratology 8:309–316 (1973).
64. Hartley J, Brown NA. Homeotic transformation in the mouse axial skeleton: a dimethadione-induced model. Teratology 50:28–29A (1994).
65. Kessel M, Gruss P. Homeotic transformations of murine vertebrae and concomitant alteration of *Hox* codes induced by retinoic acid. Cell 67:89–104 (1991).
66. Heindel JJ, Price CJ, Field EA, Marr MC, Myers CB, Morrissey RE, Schwetz BA. Developmental toxicity of boric acid in mice and rats. Fundam Appl Toxicol 18:266–277 (1992).
67. Narotsky MG, Hamby BT, Kavlock RJ. Effects of boric acid on axial skeletal development in rats. Teratology 51:192 (1995).
68. Cherrington JW, Rosen MB, Branch S, Dauterman WC, Chernoff N. Effects of boric acid administered on single days of gestation in the CD-1 mouse. Teratology 51:181 (1995).
69. Rogers JM, Mole ML, Chernoff N, Barbee BD, Turner CI, Logsdon TR, Kavlock RJ. The developmental toxicity of inhaled methanol in the CD-1 mouse, with quantitative dose–response modeling for estimation of benchmark doses. Teratology 47:175–188 (1993).
70. Connelly LE, Rogers JM. Methanol causes posteriorization of cervical vertebrae in the mouse fetus. Teratology 49: 393 (1994).
71. Bryant SV, Gardiner DM. Retinoic acid, local cell–cell interactions, and pattern formation in vertebrate limbs. Dev Biol 152:1–25 (1992).
72. Tickle C, Eichele G. Vertebrate limb development. Annu Rev Cell Biol 10:121–152 (1994).
73. Morgan BA, Tabin C. *Hox* genes and growth: early and late roles in limb bud morphogenesis. Development Suppl 181–186 (1994).
74. Doboule D. How to make a limb. Science 226:575–576 (1994).
75. Muneoka K, Anderson R. Vertebrate limb development. In: Drug Toxicity in Embryonic Development (Kavlock RJ, Daston GP, ed). Heidelberg:Springer Verlag, in press.
76. Saunders JW, Gasseling M. Ectodermal-mesenchymal interaction in the origin of limb symmetry. In: Epithelial-Mesenchymal Interaction (Flesichmayer R, Billinghman RE, eds). Baltimore:Williams and Wilkins, 1968;78–97.
77. Riddle RD, Johnson RL, Laufer E, Tabin C. *Sonic hedgehog* mediates the polarizing activity of the ZPA. Cell 75:1401–1416 (1993).
78. Tabin CJ. Why we have (only) five fingers per hand: *Hox* genes and the evolution of paired limbs. Development 116:289–296 (1992).
79. Tabin CJ. Retinoids, homeoboxes, and growth factors: toward molecular models for limb development. Cell 66:199–217 (1991).
80. Parr BA, McMahon AP. Dorsalizing signal *Wnt-7a* required for normal polarity of D-V and A-P axes of mouse limb. Nature 374:350–353 (1995).
81. Turing A. The chemical basis of morphogenesis. Philos Trans R Soc of Lond B Biol Sci 237:37–72 (1952).
82. Wolpert L. Positional information and the spatial pattern of cellular differentiation. J Theor Biol 25:1–47 (1969).
83. Wolpert L. Positional information revisited. Development Suppl 3–12 (1989).
84. Othmer HG, Maini PK, Murray JD (eds). Experimental and Theoretical Advances in Biological Pattern Formation. New York:Plenum Press, 1993.
85. Levin M. A Julia set model of field-directed morphogenesis: developmental biology and artificial life. Comput Appl Biosci 10:85–103 (1994).
86. Meinhardt H. Models of Biological Pattern Formation. London:Academic Press, 1982.
87. Oster GF, Murray JD. Pattern formation models and developmental constraints. J Exp Zool 251:186–202 (1989).
88. Murray JD, Myerscough MR. Pigmentation pattern formation in snakes. J Theor Biol 149:339–360 (1991).