Mutation Research 726 (2011) 91–97

Contents lists available at ScienceDirect

Mutation Research/Genetic Toxicology and Environmental Mutagenesis

journal homepage: www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

# Evidence for hormesis in mutagenicity dose–response relationships

Edward J. Calabrese[a,*], Edward J. Stanek III[b,1], Marc A. Nascarella[c,2]

[a] Department of Public Health, Environmental Health Sciences, University of Massachusetts, Amherst, MA 01003, United States
[b] Department of Public Health, Biostatistics and Epidemiology, Arnold House, University of Massachusetts, Amherst, MA 01003, United States
[c] Gradient, 20 University Road, Cambridge, MA 02138, United States

ARTICLE INFO

*Article history:*
Received 24 January 2011
Received in revised form 7 April 2011
Accepted 13 April 2011
Available online 21 April 2011

*Keywords:*
Mutagen
Hormetic
U-shaped
J-shaped
Dose–response
Adaptive response

ABSTRACT

This study assessed the occurrence of hormetic dose responses from three previously published data sets [1–3] with 825 chemicals in three Ames assay tester strains (i.e., TA97, TA98, TA100) with and without the S9 fraction, using a five dose protocol and semi-log dose spacing. Ninety-five (95) (11.5%) chemicals satisfied the multiple *a priori* entry criteria, with a total of 107 assays. Of the assays satisfying the entry criteria, 61 involved TA100, a strain that detects base-pair substitution mutations. 29.5% (18/61) satisfied the statistical evaluative criteria for hormesis, exceeding that predicted by chance by 4.0-fold ($p < 0.001$). The remaining 46 assays involved TA97 and TA98, strains that detect frameshift mutations. Of these 46 assays, the overall responses for the lowest two doses closely approximated the control response (e.g., 101.77% of the control for TA98; 99.20% for TA97). Only 2.2% (1/46) of the assays satisfied the evaluative criteria for hormesis. In conclusion, these data support a hormetic model for TA100, whereas the responses for TA97 and TA98 are consistent with a threshold dose–response model.



## 1. Introduction

There has long been controversy over the nature of the dose–response for carcinogens. This debate has often centered on whether the data in the low dose zone were best explained by a threshold or linear model. Since data from individual experiments have typically not been sufficient to resolve which model was the most appropriate from a statistical perspective, a public health protectionist philosophy has been adopted by advisory bodies and/or regulatory agencies, leading to the acceptance of linearity at low dose as a matter of policy. This conceptual approach was first adopted over 50 years ago by the Biological Effects of Atomic Radiation (BEAR) I Committee [4] concerning radiation-induced mutation in reproductive cells. Soon after the BEAR I report, the National Council for Radiation Protection and Measurement (NCRPM) [5] generalized the linearity-at-low-dose concept and applied it to somatic effects of mutations induced by ionizing radiation, leading to linearity at low dose modeling for the carcinogenic effects. This perspective came to be widely accepted, generalized to chemical carcinogens [6–8] and eventually integrated within the risk assessment practices of regulatory agencies throughout the world, where it is currently the dominant perspective.

The assumption that the dose of a mutagen is linearly related to response at low dose has been a central theorem underlying the regulatory approach for carcinogens.

Nevertheless, there is an increasing literature over the past decade that a hormesis model better fits response at low doses [9–18]. While the hormesis dose–response model received little attention in the 20th century [19–24], the more recent literature has demonstrated its occurrence, widespread generalizability, reproducibility, mechanistic foundations and frequency [13,14,25–28]. Several large-scale investigations have provided evidence to support claims that the hormetic dose–response model is more common than other dose–response models [29–34]. Given this resurgence of the hormetic hypothesis and its implications, we have investigated the extent to which a hormesis model is applicable in assays which detect base pair and frameshift mutations. In order to do so, we evaluated three previously published datasets which are comprised of assays for 825 chemicals utilizing five bacterial strains tested within the Ames assay, with and without rat or hamster hepatic S9 fraction activation, with a five dose framework, using semi-log dose spacing [1–3].

## 2. Methods

Mortelmans et al. [1] and Zeiger et al. [2,3] published results of mutagenicity studies with *Salmonella typhimurium* with 825 chemicals as performed by three independent laboratories. Their investigations used a suite of Ames test strains (i.e., TA97, TA98, TA100, TA1535, TA1537) with and without rat and hamster hepatic

* Corresponding author. Tel.: +1 413 545 3164; fax: +1 413 545 4692.
*E-mail addresses:* edwardc@schoolph.umass.edu (E.J. Calabrese), stanek@schoolph.umass.edu (E.J. Stanek III), mnascarella@gradientcorp.com (M.A. Nascarella).
[1] Tel.: +1 413 545 3812; fax: +1 413 545 1645.
[2] Tel.: +1 617 395 5000; fax: +1 617 395 5001.

1383-5718/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.mrgentox.2011.04.006

S9. While essentially all chemicals were tested in strains TA98, TA100, TA1535 and TA1537, only five of the 270 chemicals were tested in strain TA97 by Mortelmans et al. [1], but the strain was included in the assessments by Zeiger et al. [2,3]. In the present evaluation, only data for tester strains TA97, TA98 and TA100 are considered. Tester strains TA1535 and TA1538 are excluded due to very low background control colony counts (about 6–18 colonies per plate), resulting in high control group variability. Among the tester strains used, TA97 had a control revertant count (colonies per plate) of about 100–130, whereas TA98 and TA100 had approximately 20–35 and 90–150, respectively.

In general, an assay consists of five doses assessed per chemical with three plates per dose along with concurrent solvent (i.e., water, DMSO, 95% ethanol or acetone) controls. Available dose–response data for each assay corresponded to the average numbers of revertant colonies per plate, based on the mean of the three plates at each dose, and the standard error of the mean (SEM). Although the assays were replicated, only the average response per dose for the final replication was typically published for the individual chemicals (Zeiger, personal communication). When the results of the replicate assay did not agree or the replication was equivocal/weak, data from such assays were published along with the final "replication".

An examination of the replicated assays indicated that they were not generally designed to be exact replications. For example, in the cases of allyl isocyanate (AI) and dimethyl hydrogen phosphite (DHP), the replications did not include the lower doses (i.e., the lowest dose for AI and the two lowest doses for DHP). For methoxychlor and methdilazine, the replicates were not tested at the same percentage of S9 fraction (10% S9 in one replicate but 30% S9 in the other). In the case of methyl phenidate, the replication did not include the same tester strain (i.e., TA97) [1]. Due to the lack of exact replication, each assay was treated as an independent evaluation. Judgments were made by the original authors as to whether the agent exhibited evidence of mutagenicity. In their papers [1–3], an agent was deemed to cause a mutagenic response if responses were dose related, causing a reproducible increase in the number of revertants above background, even if the increase was less than two-fold.

### 2.1. Entry criteria

The present paper assessed the frequency of hormetic dose responses for mutagenic endpoints within the Ames assay, using the data sets of Mortelmans et al. [1] and Zeiger et al. [2,3]. To assess the frequency of hormesis, *a priori* entry and evaluative criteria were employed as described below (Fig. 1).

#### 2.1.1. Entry Criteria #1: response at dose #5 is ≥110% of control and entry Criteria #2: response at dose #4 is <110% of control

The initial entry criteria (Criteria #1 and #2) involved the selection of dose responses which had a local Benchmark Dose $(BMD)_{10}$ [31]. The local $BMD_{10}$ is defined as a dose between dose #4 and dose #5 such that the response at dose #4 is below 110% and that at dose #5 is ≥110% of the control. With particular reference to the data we describe, this means that there are >10% more mutations (i.e., increase in the number of revertants) at Dose 5 relative to the control. This ensures that a response at dose #5 would equal or exceed 110% of the control rate and thereby provide potential evidence of a mutagenic response. This tentative conclusion is further strengthened by statistical evaluation (see Criteria #5). These criteria resulted in assays having three doses below the lower bound (LB) of the local $BMD_{10}$.

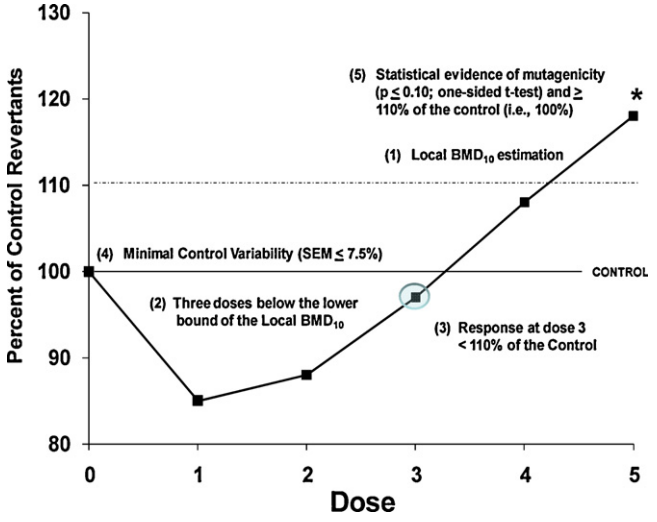


**Fig. 1.** Dose–response entry criteria. (1) The evaluative strategy centered on assessing the responses of the two lowest doses (doses #1 and #2) below the lower bound (LB) dose of the $BMD_{10}$. Dose #5 was required to satisfy two minimum criteria statistic evidence of mutagenicity ($p \leq 0.1$), one-side *t*-test and have a response ≥110% of the control group (100%). (2) The LB dose bracketing the $BMD_{10}$ (i.e., dose #4) and the third lowest dose (i.e., dose #3) were not used to assess hormesis. Responses for the doses #3 and #4 were required to be <110%. (3) In an effort to minimize variability, a criterion also required that the control group of each dose–response display an SEM of <7.5%. (4) A dose–response demonstrated evidence of hormesis if one or two mean responses for doses #1 and #2 were significantly less than the control response (two-way *t*-test, $p < 0.10$).

#### 2.1.2. Entry Criteria #3: response to dose #3 is <110% of the control

In order to enhance the likelihood of a more stable and accurate estimate of the toxicological threshold, dose responses were eliminated if the response of the third highest dose (i.e. dose #3) was ≥110% of the control.

#### 2.1.3. Entry Criteria #4: minimize control group variation via SEM limit

The fourth entry criteria required that the control group of each dose–response display a standard error of the mean (SEM) of ≤7.5% to minimize variability. We used this criterion to select assays with higher power to detect differences in response at low doses.

#### 2.1.4. Entry Criteria #5: statistical evidence of mutagenicity at dose #5

Entry criteria #5 required that the *p*-value for a one-sided test of the null hypothesis that response at dose #5 is less than or equal to 110% of control be rejected with a *p* value of ≤0.10.

#### 2.1.5. Simulation study validation of entry criteria and tester strain hyper-poisson control group distribution assumption

Detailed simulation studies assessed the impact of each *a priori* entry criteria specifically or in combination. These simulations used several different strategies. One involved an assumption of a normal distribution of control counts and employed an estimate of a standard deviation to generate simulated individual "experiments" to assess the role of bias in selection criteria. A second approach involved the use of data simulated directly from repeated measures or control samples [35]. This second approach was designed to account for the possible hyper-poisson distribution that has been proposed for control samples, thereby providing a sensitivity analysis to the normality assumption. The results from both simulation strategies revealed no evidence that bias was introduced into the assessment of hormetic responses at low doses.

### 2.2. Evaluative strategy

#### 2.2.1. Evaluative criteria for low potency mutagens

The evaluative strategy centered on assessing the responses of the two lowest doses (doses #1 and #2) below the lower bound (LB) dose of the $BMD_{10}$. The lower bounding dose bracketing the $BMD_{10}$ (i.e., dose #4) and the third lowest dose (i.e., dose #3) were not used to assess the possible occurrence of hormesis. An assay was considered to provide evidence of hormesis if a two sided test of equality of response to control was rejected at $\alpha = 0.10$ at dose #1 and/or dose #2. A similar assessment was also made using $\alpha = 0.05$.

We classified assays with a statistically significant result as having response below the control (which we refer to as hormesis) or response above the control. We based the classification on doses where the response was statistically significant. In no case was there a conflict in the direction of significance.

#### 2.2.2. Type I error—false positive estimation

Associated with any hypothesis test is the type I error, equal to the probability of rejecting the null hypothesis when the null is true. Such rejected tests are false positive results. Since two statistical tests were conducted for each assay, the probability that one (or more) test is statistically significant by chance is given by $1-(1-\alpha)^2$, i.e., 0.19 when $\alpha = 0.10$ and 0.0975 when $\alpha = 0.05$. We use these false positive rates to predict the number of false positive assays, which we expect to be evenly distributed above and below the control mean.

#### 2.2.3. Additional statistical criteria

We were particularly interested in whether or not the response was above or below control response when the difference was considered to be statistically significant. To answer this question, we tested the null hypothesis that the proportion of statistically significant assays above and below control was equal using a binomial test.

#### 2.2.4. Entry/evaluative criteria for chemicals with higher mutation production

*Entry criteria*: Criteria #1 – The response of dose 5 must be ≥150% relative to the concurrent control of 100% and statistically significantly greater than the control ($p < 0.1$). Criteria #2 – The dose–response must display a monotonic decrease in response from dose 5 to dose 4 to dose 3. Criteria #3 – The response of dose 3 must be ≤110% compared to the control of 100%. The evaluative criteria are the same as for the mutagens described in Section 2.2.1.