PPPTP TESTING AND ANALYSIS METHODS

In vitro testing: *In vitro* testing was performed using DIMSCAN, a semiautomatic fluorescence-based digital image microscopy system that quantifies viable (using fluorescein diacetate [FDA]) cell numbers in tissue culture multiwell plates [2]. Cells were incubated in the presence of drug for 96 hours. Mean fluorescence values were determined for each concentration tested and then normalized to the mean control fluorescence for the line to determine relative mean fluorescence values.

For analysis of *in vitro* testing results, a non-linear regression, sigmoidal dose-response model was fitted using GraphPad Prism 5.03 to the relative mean fluorescence values *vs*. the log-transformed concentration (X) for the *in vitro* PPTP study data:

$$Y=Bottom + (Top-Bottom)/(1+10^{((LogrIC_{50}-X)*HillSlope)})$$

The terms are defined as follows: rIC_{50} (relative IC_{50}) is the concentration of agent that gives a response half way between Bottom and Top; HillSlope describes the steepness of the dose-response curve; and Top and Bottom are the plateaus in the T/C% values at low and high concentrations, respectively. Absolute IC_{50} values represent the concentration at which the agent reduces cell survival to 50% of the control value [3]. To compare activity between cell lines, the ratio of the median relative IC_{50} to individual cell line's relative IC_{50} value is used (larger values connote greater sensitivity). Observed Y_{min} is the minimum survival fraction (Treated/Control %) at the range of concentrations of the drug employed.

The Relative In/Out (I/O)% values compare the relative difference in final cell number compared with the starting cell number for treated cells and for control cells calculated as follows: (Observed $Y_{min}-Y_0$)/(100−Y_0) if Observed $Y_{min}>Y_0$; and (Observed $Y_{min}-Y_0$)/(Y_0) if Observed Y_{min}<Predicted Y_{min}). Y_0 is an estimate of the starting cell number derived from determinations of the doubling time of the cell line. Relative I/O% values range between 100% (no treatment effect) to -100% (complete cytotoxic effect), with a Relative I/O% value of 0 being observed for a completely effective cytostatic agent.

The F test was used to determine whether there is statistical evidence for a plateau at higher concentrations. The null hypothesis is that the simpler model (bottom T/C% = 0) is correct and the alternative is that not constraining the bottom T/C% value is correct. If the F test P value is < 0.05, then the conclusion is that the more complicated model with the bottom unconstrained fits significantly better and that the agent therefore has a non-zero plateau effect at higher concentrations.

In vivo testing: CB-17SC female mice (Taconic Farms, Germantown NY), were used to propagate subcutaneously implanted kidney/rhabdoid tumors, sarcomas (Ewing, osteosarcoma, rhabdomyosarcoma), neuroblastoma, and non-glioblastoma brain tumors, while BALB/c nu/nu mice were used for glioma models, as previously described [4-6]. Mice bearing subcutaneous tumors each received drug when tumors reached between 200 mm^3 and 500 mm^3. Human leukemia cells were propagated by intravenous inoculation in female non-obese diabetic (NOD)/*scid*$^{-/-}$ mice as described previously [7]. Mice were randomized to groups of 10 for solid tumor-bearing and groups of 8 for ALL-bearing animals. All mice were maintained under barrier conditions and experiments were conducted using protocols and conditions approved by the institutional animal care and

use committee of the appropriate consortium member. Each agent tested was given a code number, and the identity of each was revealed to testing sites only after complete data sets had been deposited in the database.

Tumor volumes (cm^3) [solid tumor xenografts] or percentages of human CD45-positive [hCD45] cells [ALL xenografts] were measured for each tumor at the initiation of the study and weekly for up to 42 days after study initiation. Assuming tumors to be spherical, tumor volumes were calculated from the formula $(\pi/6)\cdot d^3$, where d represents the mean diameter.

Response and Event Definitions for Solid Tumor Xenograft Models

Response: For individual mice, progressive disease (PD) was defined as < 50% regression from initial volume during the study period and > 25% increase in initial volume at the end of study period. Stable disease (SD) was defined as < 50% regression from initial volume during the study period and ≤ 25% increase in initial volume at the end of the study. Partial response (PR) was defined as a tumor volume regression ≥50% for at least one time point but with measurable tumor ($\geq$ 0.10 cm^3). Complete response (CR) was defined as a disappearance of measurable tumor mass (< 0.10 cm^3) for at least one time point. A complete response was considered maintained (MCR) if the tumor volume was <0.10 cm^3 at the end of the study period. For treatment groups only, if the tumor response was PD, then PD was further classified into PD1 or PD2 based on the tumor growth delay (TGD) value. TGD values were calculated based on the numbers of days to event. For each individual mouse that had PD and had an event in the treatment groups, a TGD value was calculated by dividing the time to event for that mouse by the median time to event in the respective control group. Median times to event were estimated based on the Kaplan-Meier event-free survival distribution. If a mouse had a TGD value ≤ 1.5, that mouse was considered PD1. If the TGD value was > 1.5, the mouse was considered PD2. Mice that had PD but did not have an event at the end of the study were coded as PD2.

Event-free survival: An event in the solid tumor xenograft models was defined as a quadrupling of tumor volume from the initial tumor volume. Event-free survival was defined as the time interval from initiation of study to the first event or to the end of the study period for tumors that did not quadruple in volume. The time to event was determined using interpolation based on the formula:

$$t_x = t_1 + (t_2 - t_1)\ln(V_e / V_1) / \ln(V_2 / V_1),$$

where t_x is the interpolated day to event, t_1 is the lower observation day bracketing the event, t_2 is the upper observation day bracketing the event, V_1 is the tumor volume (or hCD45 percentage) on day t_1, V_2 is the tumor volume (or hCD45 percentage) on day t_2 and V_e is the event threshold (4 times initial tumor volume for solid tumor xenografts, 25% for ALL xenografts).

Response and Event Definitions for Acute Lymphoblastic Leukemia (ALL) Xenograft Models

Individual mice were categorized as PD if their percentage of hCD45 cells never dropped below 1% and they had an event before the end of the study period. An event is defined as hCD45 cells above 25% in the peripheral blood with times to event calculated as above. Individual mice were classified as SD if their percentage of hCD45 cells never dropped below 1% and no event occurred before the end of the study. PR was assigned if the percentage of cells dropped below 1% for any one time point regardless of whether the

percentage eventually reached 25%. A CR was assigned if the percentage of hCD45 cells dropped below 1% for 2 consecutive weeks of the study and regardless of whether the percentage reached 25% or not. A CR was considered maintained if the percentage of hCD45 was less than 1% for the last three measurements of the study. For treatment groups, PD was further classified into PD1 and PD2 according to the TGD value.

Summary statistics and analysis methods

Overall Group Response: Each individual mouse was assigned a score from 0 to 10 based on their response: PD1=0, PD2=2, SD=4, PR=6, CR=8, and MCR=10, and the median for the group determined the overall response. Studies in which toxicity was greater than 25% or in which the control group was not at least SD, were considered inevaluable and were excluded from analysis. Treatment groups with PR, CR, or MCR are considered to have had an objective response. Agents inducing objective responses are considered highly active against the tested line, while agents inducing stable disease or PD2 are considered to have intermediate activity, and agents producing PD1 are considered to have a low level of activity against the tested line.

Tumor Volume T/C value: Relative tumor volumes (RTV) for control (C) and treatment (T) mice were calculated at day 21 or when all mice in the control and treated groups still had measurable tumor volumes (if less than 21 days). The mean relative tumor volumes for control and treatment mice for each study were then calculated and the T/C value was the mean RTV for the treatment group divided by the mean RTV for the control group. For the tumor volume T/C response measure, agents producing a T/C of $\leq$ 15% are considered highly active, those with a mean tumor volume T/C of $\leq$ 45% but > 15% are considered to have intermediate activity, and those with mean T/C values > 45% are considered to have low levels of activity [8].

EFS T/C value: An EFS T/C value was defined by the ratio of the median time to event of the treatment group and the median time to event of the respective control group. If the treatment group did not have a median time to event, then EFS T/C was defined as greater than the ratio of the last day of the study for the treatment group divided by the median time to event for the control group. For the EFS T/C measure, agents are considered highly active if they meet three criteria: a) an EFS T/C > 2; b) a significant difference in EFS distributions ($p \leq 0.050$), and c) a net reduction in median tumor volume for animals in the treated group at the end of treatment as compared to at treatment initiation. Agents meeting the first two criteria, but not having a net reduction in median tumor volume for treated animals at the end of the study are considered to have intermediate activity. Agents with an EFS T/C < 2 are considered to have low levels of activity. Xenografts in which the median EFS for the control line was greater than one-half of the study period or in which the median EFS for the control line did not exist are considered not evaluable for the EFS T/C measure of activity.

Statistical Methods: The exact log-rank test, as implemented using Proc StatXact for SAS®, was used to compare event-free survival distributions between treatment and control groups. P-values were two-sided and were not adjusted for multiple comparisons given the exploratory nature of the studies. The exact Wilcoxon rank-sum test was used to test the difference of medians of EC_{50} values between the groups of lines with similar tumor types to the remaining lines of the panel. Fisher's exact test and the two-sample t-test, respectively, were used to compare mortality rates and average lowest weights between treated and control groups. The Mann-Whitney test was used to test the difference of the median values of the cytotoxic activity for cell lines of individual panels compared to the remaining PPTP cell lines.

REFERENCES

1. Kang MH, Smith MA, Morton CL, et al. National Cancer Institute Pediatric Preclinical Testing Program: Model description for in vitro cytotoxicity testing. Pediatr Blood Cancer 2011:56(2):239-249.
2. Frgala T, Kalous O, Proffitt RT, Reynolds CP: A novel cytotoxicity assay with a 4 log dynamic range that identifies synergistic drug combinations. Molecular Cancer Therapeutics 2007:6:886-89.
3. Keshelava N, Frgala T, Krejsa J, et al. DIMSCAN: a microcomputer fluorescence-based cytotoxicity assay for preclinical testing of combination chemotherapy. Methods in molecular medicine 2005:110:139-153.
4. Sebaugh JL. Guidelines for accurate EC50/IC50 estimation. Pharmaceut Statist 2011:10 128–134.
5. Friedman HS, Colvin OM, Skapek SX, et al. Experimental chemotherapy of human medulloblastoma cell lines and transplantable xenografts with bifunctional alkylating agents. Cancer research 1988:48(15):4189-4195.
6. Graham C, Tucker C, Creech J, et al. Evaluation of the antitumor efficacy, pharmacokinetics, and pharmacodynamics of the histone deacetylase inhibitor depsipeptide in childhood cancer models in vivo. Clin Cancer Res 2006:12(1):223-234.
7. Peterson JK, Tucker C, Favours E, et al. In vivo evaluation of ixabepilone (BMS247550), a novel epothilone B derivative, against pediatric cancer models. Clin Cancer Res 2005:11(19 Pt 1):6950-6958.
8. Liem NL, Papa RA, Milross CG, et al. Characterization of childhood acute lymphoblastic leukemia xenograft models for the preclinical evaluation of new therapies. Blood 2004:103(10):3905-3914.
9. Plowman J, Camalier R, Alley M, et al. US-NCI testing procedures. Relevance of tumor models for anticancer drug development. Basel: Karger; 1999. p 121-135.