AIR Binder: An Automatic Reporting and Data Analysis SAS® Application for Cytochrome P450 Inhibition Assay to Investigate DDI

Hao Sun, Covance Inc., Madison, Wisconsin, United States
Kristen Cardinal, Covance Inc., Madison, Wisconsin, United States
Carole Kirby, Covance Inc., Leeds, United Kingdom
Richard Voorman, Covance Inc., Madison, Wisconsin, United States

ABSTRACT

We developed a SAS®-based application, called AIR Binder (Automatic Inhibition Reporting Binder), which integrates a series of SAS® macros to perform data analysis and visualization, and generate report-ready tables and figures for the cytochrome P450 (CYP) inhibition assay, a widely-used in vitro ADME assay that provides key supporting data for drug discovery and development. The dataset from this assay was designed to have only 3 columns (assay/sample identifier, concentration, and activity) to simplify data management/storage and QC process. AIR Binder has been successfully used to analyze this type of dataset, generate required statistics, make report-ready comprehensive tables, and plot high-quality grouped figures. The program calculates IC_{50} values using non-linear fitting algorithms based on several variations of a 4-parameter logistic (4PL) model. AIR Binder is a standardized and customized program for efficiently reporting CYP inhibition data, significantly reducing turnaround time, eliminating unnecessary errors due to data transfer, minimizing labor cost, and enhancing productivity.

INTRODUCTION

CYP inhibition is a standardized in vitro ADME assay that runs on an automated liquid handling system in most pharmaceutical companies and CROs. Relatively high-throughput bioanalysis often via LC/MS is followed by data analysis to generate summarized enzyme "activity" results. A traditional way of data analysis, table making, figure plotting, and curve fitting may involve various software packages. For example, in our organization Microsoft Excel was used to make tables and plot figures and Systat SigmaPlot or GraphPad Prism was used for non-linear curve fitting due to their integrated kinetic equations. However, data generated from a CYP inhibition assay for a single compound will require compilation of approximately 30-40 tables, 70-80 figures, and 20-30 IC_{50} calculations and fitting curves, which makes the whole process extremely time-consuming and inefficient. In addition, due to data transfer between different software packages, extensive QC was required. Although macros can be incorporated into these software packages, overall productivity was still limited. If multiple compounds are involved for a single study, the numbers of tables and figures can become overwhelming. Therefore, we developed a novel solution, AIR Binder, using SAS® as a platform to handle data compiling, analysis, and visualization, for table and figure generation and IC_{50} calculation with curve fitting. From data input to report-ready table and figure output, there is no need for any intermediate processing, and the whole process is accomplished instantly.

AIR BINDER WORKFLOW

The overall workflow of AIR Binder is described in Figure 1. Specifically, the input to AIR Binder and the output to tables and figures is a simple two-step process. Input tables from the CYP inhibition assays include 3 columns: assay/sample identifier, compound concentration, and enzyme activity. Several SAS® procedures were applied for the overall process. The MEANS, FORMAT, TRANSPOSE, and REPORT procedures were used for data compiling and table making. The SGPANEL and SGPLOT procedures were used for plotting figures. The NLIN procedure was implemented for non-linear curve fitting using a 4-parameter logistic model. All functions were packaged into SAS® macros and can be called just once to generate an overall report.
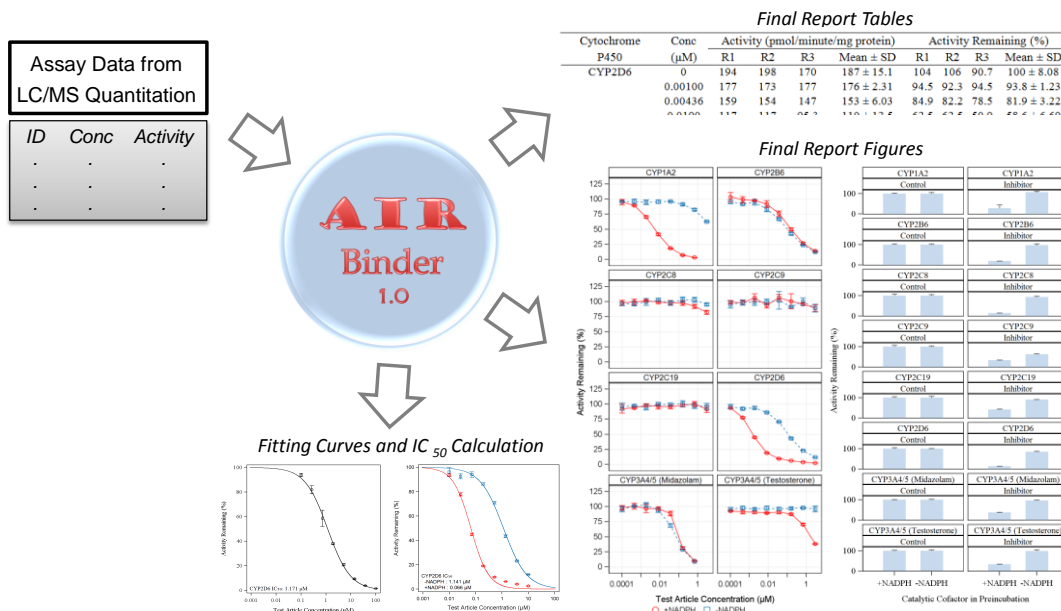
Figure 1. Workflow of AIR Binder

DATA COMPILING AND PROCESSING

The overall CYP inhibition assay contains two separate experiments, direct inhibition and metabolism-dependent inhibition, designed as key components for the prediction of bioavailability, drug-drug interactions and toxicity. In general, each assay tests 7-8 major hepatic CYP isoforms using human liver microsomes in the presence of specific marker substrates. Unique positive controls for each isoform are included for comparison. Data structures are the same for all assays (for both test article and controls), and the unique identifier column differentiates assay conditions. An external text file was implemented in AIR Binder as a lookup table to format assay conditions such as test article vs. control, specific CYP isoforms tested, and with/without catalytic cofactors in the preincubation. PROC FORMAT with the CNTLIN option was used to read in this external text file (the "infile" parameter) and to build an ordered format (the "fmtname" parameter) for table columns and figure labels, as follows:

```
%macro inhfmt(infile=infile,fmtname=in);
   data _fmt;
       infile &infile dsd dlm=',';
       retain fmtname &fmtname type 'n';
       input start label :$50.;
   run;
   proc format cntlin=_fmt;
   run;
%mend;
```

The external text file is convenient for maintenance when assay condition needs to be updated. In addition, this specific technique is efficient to order CYP names. For 57 human CYPs identified to date, they were named by family (CYP1,2,3...), subfamily (A,B,C...) and member order (1,2,3...), a mixed string and number naming system. For example, major human CYPs routinely tested were CYP1A2, 2B6, 2C8, 2C9, 2C19, 2D6, and 3A4/5. Without formatting, CYP2C19 would appear ahead of CYP2C8 unless additional text processing is applied. Thus, PROC FORMAT is an easy one-step solution.

The mean and standard deviation of enzyme activities were calculated by PROC MEANS. The percent of enzyme activity remaining was normalized by solvent controls (activity when test article concentration is 0). Similarly, the mean and standard deviation of the percent enzyme activity remaining were also calculated by PROC MEANS.

Rounding numbers to the required significant figures is a key step for data processing. The actual "observation" from the CYP inhibition assay was the activity column, which should be rounded to 3 significant figures per Covance SOP. The dependent variables and statistics such as mean, standard deviation, and percent activity remaining were calculated in AIR Binder based on the rounded 3-significant-figure activity values. In this way, the whole table can be easily verified by the client and/or agency for consistency. A function to round variables to their n-significant figure format was developed ("roundstat", as follows). One or multiple variables (separated by a single white space) can be defined with its "roundvarlist" parameter. The CALL SYMPUT routine was used to extract real variable names from the list, and the &&VAR&I macro variable was used to round variables with the ROUND function. Missing data values due to instrument errors were assigned as "NA" in tables; data values below bioanalytical detection limit was assigned as "0".

```
%macro roundstat(infile=in,roundvarlist=in,outfile=out,sigfig=3);

   %let dimvar=0;
   %let outprefix=temp;
   %do %while(%qscan(&roundvarlist,&dimvar+1,%str( )) ne %str( ));
       %let dimvar=%eval(&dimvar+1);
   %end;
   %do s=1 %to &dimvar;
       %let varname=%scan(&roundvarlist,&s,' ');
       data _null_;
             call symput('var'||left(&s),"&varname");
       run;
   %end;
   data &outprefix.0;
       set &infile;
   run;
   %let j=0;
   %do i=1 %to &dimvar;
      data &outprefix&i;
       set &outprefix&j;
       if int(&&var&i) ne 0 then do;
       &&var&i=round(&&var&i,10**(int(log10(abs(&&var&i)))-&sigfig.+1));
       end;
       if int(&&var&i) eq 0 and &&var&i ne 0 then do;
       &&var&i=round(&&var&i,10**(-1*(abs(int(log10(abs(&&var&i))))+&sigfig.)));
       end;
       run;
       %let j=%eval(&j+1);
       %put Rounding &&var&i Completion!;
   %end;
   data &outfile;
       set &outprefix&dimvar;
   run;
%mend;
```

TABLE DESIGN AND LAYOUT

Tables for the CYP inhibition assay were produced with the REPORT procedure. In general, four separate tables were produced per test article per study, with each table containing all CYP isoforms tested. The identifier, concentration, activity, and calculated statistics mentioned previously were combined for an overall table presentation. Replicates at each concentration were transposed into one line using the TRANSPOSE procedure, which removed all unnecessary white space within tables for a concise design. Another improved feature of the tables was the combination of the mean and standard deviation columns. The CATX function was used to merge them with a "±" sign added in between. Together, all assay tables were condensed in only a few pages, which made it easier to review and compare by lab scientists, clients, and agency (Figure 2).

	STYLE	Cytochrome	Conc	Activity (pmol/minute/mg protein)				Activity Remaining (%)			
STYLE(HEADER) ⇨		P450	(μM)	R1	R2	R3	Mean ± SD	R1	R2	R3	Mean ± SD
		CYP3A4/5	0	3270	3640	3520	3480 ± 189	94.1	105	101	100 ± 5.43
		(Testosterone)	0.000100	3290	3190	3230	3240 ± 50.3	94.6	91.8	92.9	93.1 ± 1.45
SPANROWS ⇨		+NADPH	0.000436	3050	3210	3290	3180 ± 122	87.7	92.3	94.6	91.6 ± 3.51
			0.00190	3210	3210	3070	3160 ± 80.8	92.3	92.3	88.3	91.0 ± 2.32
			0.00829	3150	3050	3130	3110 ± 52.9	90.6	87.7	90.0	89.5 ± 1.52
			0.0362	3230	3190	3050	3160 ± 94.5	92.9	91.8	87.7	90.8 ± 2.72
			0.158	3010	3110	3010	3040 ± 57.7	86.6	89.5	86.6	87.5 ± 1.66
			0.688	2520	2410	2430	2450 ± 58.6	72.5	69.3	69.9	70.6 ± 1.69
CALL DEFINE ⇨			3.00	1300	1340	1340	1330 ± 23.1	37.4	38.5	38.5	38.2 ± 0.664
		CYP3A4/5	0	3620	3270	3330	3410 ± 187	106	96.0	97.7	100 ± 5.49
		(Testosterone)	0.000100	3170	3330	3310	3270 ± 87.2	93.1	97.7	97.2	96.0 ± 2.56
		-NADPH	0.000436	3370	3410	3230	3340 ± 94.5	98.9	100	94.8	97.9 ± 2.77
SPANROWS ⇨			0.00190	3290	3190	3310	3260 ± 64.3	96.6	93.6	97.2	95.8 ± 1.89
			0.00829	3420	3210	3370	3330 ± 110	100	94.2	98.9	97.8 ± 3.22
			0.0362	3190	3290	3350	3280 ± 80.8	93.6	96.6	98.3	96.2 ± 2.37
			0.158	3250	3350	3310	3300 ± 50.3	95.4	98.3	97.2	97.0 ± 1.48
			0.688	3350	3290	3370	3340 ± 41.6	98.3	96.6	98.9	97.9 ± 1.22
CALL DEFINE ⇨			3.00	3460	3130	3250	3280 ± 167	102	91.9	95.4	96.3 ± 4.90
		⇧		⇧		⇧		⇧		⇧	
		FORMAT		**TRANSPOSE**		**CATX**		**TRANSPOSE**		**CATX**	

Figure 2. Tables generated by AIR Binder

To display rounded numbers with 3 significant figures in tables, an additional step was conducted. The function was written similarly as the function "roundstat" described previously using the CALL SYMPUT routine and &&VAR&I macro variable to format a list of columns to display. Specifically, a new macro variable "&&sf&&var&i" was used to store the converted strings with the PUT function in a W.D. format for a wide range of numbers ($10^{-6} \le N < 10^{12}$), as follows:

```
data &outprefix&i(drop=&&var&i);
    set &outprefix&j;
        if &&var&i lt 0.000001 then &&sf&&var&i= put(&&var&i,12.0);
        else if &&var&i lt 0.00001 then &&sf&&var&i=put(&&var&i,10.8);
        else if &&var&i lt 0.0001 then &&sf&&var&i=put(&&var&i,9.7);
        else if &&var&i lt 0.001 then &&sf&&var&i=put(&&var&i,8.6);
        else if &&var&i lt 0.01 then &&sf&&var&i=put(&&var&i,7.5);
        else if &&var&i lt 0.1 then &&sf&&var&i=put(&&var&i,6.4);
        else if &&var&i lt 1 then &&sf&&var&i=put(&&var&i,5.3);
        else if &&var&i lt 10 then &&sf&&var&i=put(&&var&i,4.2);
        else if &&var&i lt 100 then &&sf&&var&i=put(&&var&i,4.1);
        else &&sf&&var&i=put(&&var&i,12.0);
run;
```

After the data format was finalized, PROC REPORT was applied to define the columns to display on tables and thus draw tables. Simply, three techniques were used to add/remove lines of tables (Figure 2). First, STYLE and STYLE(HEADER) were applied to define table border color. In addition, the SPANROWS option of PROC REPORT was used to create spanning cells for the formatted and ordered identifier variable (the first column), and thus to remove lines within one spanning cell. Also, the CALL DEFINE statement with a couple of COMPUTE blocks was used to find line positions (the highest concentration tested for each identifier group). A separate black line was drawn below each identifier group, a visual optimization for big tables.

```
define &linepos / display noprint;
compute &linepos;
    if &linepos ^=. then hold=&linepos;
endcomp;
compute &x;
    if hold=&x then do;
        call define(_row_,'style','style=[borderbottomwidth=1pt
        borderbottomcolor=black]');
    end;
endcomp;
```

IC$_{50}$ CALCULATION AND NON-LINEAR CURVE FITTING

IC$_{50}$ values were calculated in AIR Binder with a 4-parameter logistic model that is decreasing, using the NLIN procedure with a general formula as follows:

$$f(x) = \Phi_0 + \frac{\Phi_1 - \Phi_0}{1 + \left(\frac{x}{\Phi_2}\right)^{-\Phi_3}}$$

where Φ_0 is the lower asymptote, Φ_1 is the upper asymptote, Φ_2 is the inflection point or IC$_{50}$, and Φ_3 is the Hill slope. The ODS OUTPUT was implemented to collect estimated model parameters from PROC NLIN, as follows. The smoothness of fitting curves was controlled by a macro variable "&iter".

```
    ...part of a 4PL preparation function...

data _onecyp&cypname;
    set &infile;
    if enzyme="&cypname";
run;
ods output parameterestimates=_list&cypname(keep=parameter estimate);
proc nlin data=_onecyp&cypname best=3 maxiter=983 method=marquardt;
    parms max=100 min=0 ic50=&ic50est hill=-1;
    model &y=&mincal+(&maxcal-&mincal)/(1+(conc/ic50)**(-hill));
run;
ods output close;
proc transpose data=_list&cypname out=_listout&cypname (drop=_name_);
    var estimate;
run;
data _null_;
    set _listout&cypname;
    call symput('inset_ic50value'||&cypnum, put(&ic50col, bestd5.3));
run;
data _newlistout&cypname (drop=&dropcol);
    set _listout&cypname;
    &x=&minconc;
    do i=0 to &iter*&maxconc;
        &curvename=&mincol+(&maxcol-&mincol)/(1+(&x/&ic50col)**(-&hillcol));
        &fitname="&cypname";
        output;
        x=&x+(1/&iter);
    end;
run;
```

Four variations of IC$_{50}$ fitting were implemented using a combination of macro variables: "&mincol", "&maxcol", "&ic50col", and "&hillcol". A relative IC$_{50}$ (or IC$_{50}$) was determined without constraining these variables. An absolute IC$_{50}$ was determined by constraining both upper and lower asymptotes. The other two variations constrain either upper or lower asymptote (Figure 3).
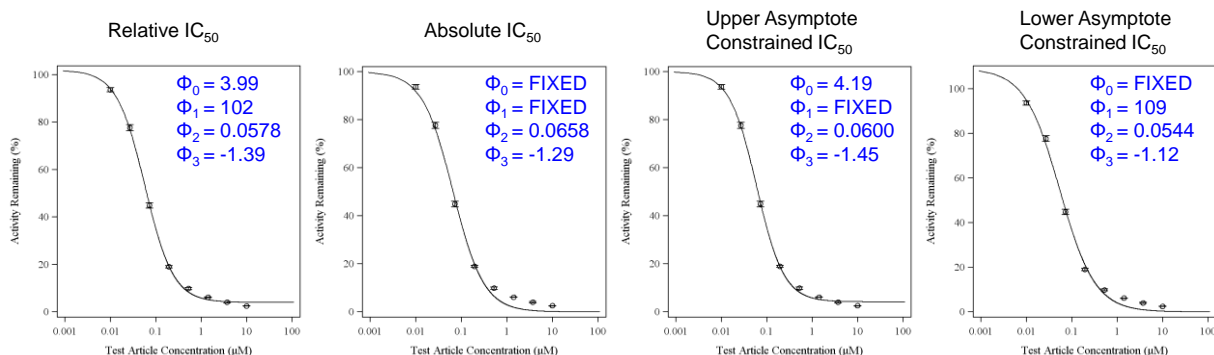

Figure 3. Variations of 4-parameter logistic IC$_{50}$ fitting

Because a normalized percentage was modeled and the lower asymptote in CYP inhibition assay tended to be zero (but may never reach zero due to limitation of solubility), a logistic equation that only estimates the upper asymptote (Φ_1), IC$_{50}$ (Φ_2) and Hill slope (Φ_3) was used more often.

A nonlinear least squares method using the Levenberg-Marquardt algorithm was implemented to fit the IC$_{50}$ curve, which regresses the residuals onto the partial derivatives of the 4-parameter logistic model until the estimates converge, with an updating formula (Δ) as,

$$\Delta = \left(X^{'}X + \lambda\, diag(X^{'}X) \right)^{-} X^{'}r$$

where λ is the combination coefficient, $X'X$, $diag(X'X)$ and $X'r$ are the Hessian, the diagonal of Hessian, and the gradient, respectively, of the residual sum of squares (SSE) as follows,

$$SSE(\beta) = \sum_{i=1}^{n}(y_i - f_i(\beta))^2 = r(\beta)^{'}r(\beta)$$

where y_i is the ith observation, $f_i(\beta)$ is the ith equation for the 4-parameter logistic function, and r is the residual vector, for a given parameter vector β.

GRAPHIC PRESENTATION

Graphic presentation was a key consideration for the interpretation of CYP inhibition data. The SCATTER and SERIES plots were implemented in AIR Binder using PROC SGPANEL with ONEPANEL display (Figure 4). A panel plot containing all CYP isoforms can demonstrate their similarity and difference instantly. For the metabolism-dependent assay, a plot overlay by different conditions also facilitates decision making. The ODS GRAPHICS statement was used to create antialiased PNG images, which were then output with the ODS RTF statement. Specifically, for the line overlay plots, the color and shape of markers and lines were modified by %MODSTYLE, a build-in SAS® style template modification macro. The CLM type was selected to simultaneously cycle through colors, markers, and line styles. In addition, the size and style of panel organization were controlled with an overall macro call, which include the height and width of output panel images and number of columns.
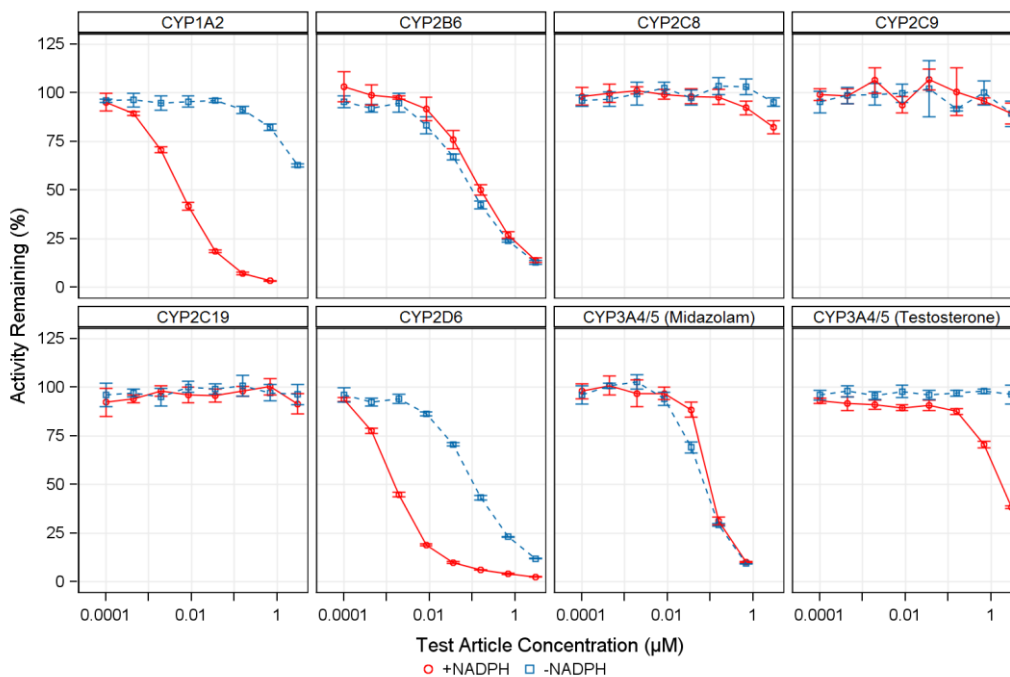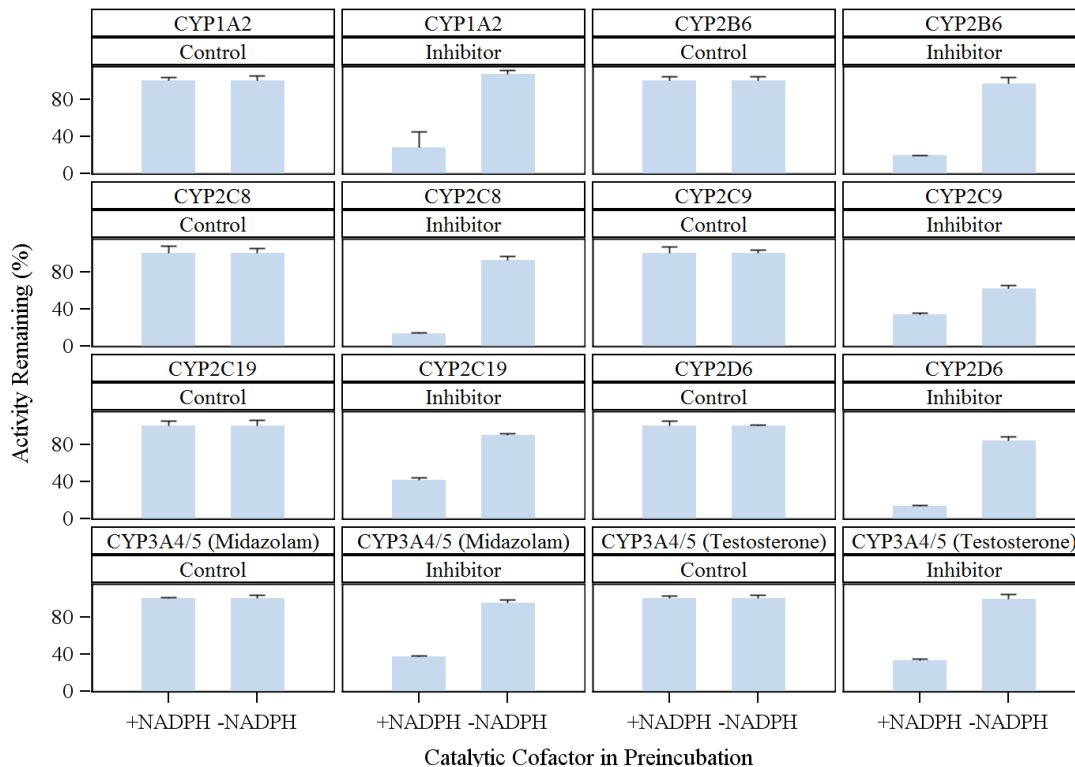

Figure 4. Panel plot of inhibition by CYP isoforms

Figure 5. Panel bar chart of positive and negative controls

For the graphic presentation of positive and negative controls, a panel bar chart using PROC SGPANEL with the VBAR statement was designed in AIR Binder (Figure 5). No full IC_{50} curves were determined for control samples; only single-concentration inhibition was determined and presented. The difference between two vertical bars (with and without the catalytic cofactor NADPH in the preincubation) is easy to observe, and thus assay quality is easy to be evaluated.

In addition, IC_{50} curves were plotted by the SGPLOT procedure, with 3 options available to call in AIR Binder, a single curve plot, a double curve overlay, or a triple curve overlay (Figure 6). IC_{50} values were designed to be present directly on these plots using the INSET statement. The IC_{50} values were saved first using a group of global macro variables ('inset_ic50value' || &cypnum) via the CALL SYMPUT routine in the 4PL preparation function, as described above.
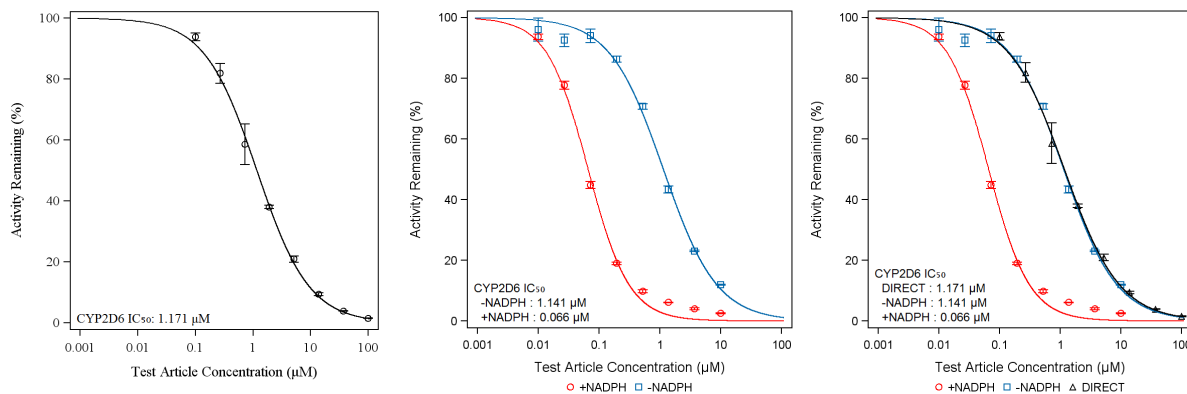

Figure 6. Curve fitting using 4PL with inset IC_{50} display

APPLICATION IN OTHER IN VITRO ADME ASSAYS

The same workflow used in AIR Binder has been extended to other in vitro ADME assay reporting, such as induction, phenotyping, metabolic stability, protein binding, and cross-species profiling and metabolite identification. For example, experimental data from induction assay has a very similar structure. Instead of determining percent activity remaining, it evaluates fold-change of activity at two levels: protein activity induced and mRNA expression elevated. A similar 4-parameter logistic model that is increasing has been used to determine EC_{50}. For some in vitro assays, curve fitting may not be needed; thus the table making and figure plotting functions in AIR Binder can be applied directly. If additional linear or nonlinear fitting method is needed, such as the determination of half-life in microsome/hepatocyte metabolic stability assays, a separate SAS® function can be written and easily integrated to the current AIR Binder workflow. Similarly, a companion assay to the CYP inhibition assay to report the K_I and k_{inact} values of a potential CYP inactivator is a component of AIR Binder, which contains a different kinetic curve fitting algorithm and data visualization scheme. Therefore, AIR Binder can be further developed as a powerful tool package for reporting all regular standardized in vitro ADME assays.

CONCLUSION

SAS® is a powerful platform for development of customized applications on automatic data analysis and reporting of in vitro/in vivo assays, together with preclinical and clinical pharmacokinetic and drug metabolism findings. AIR Binder, a specific SAS® application to analyze and report CYP inhibition assay data, demonstrated significantly enhanced productivity including reduced turnaround time, better designed tables and figures, and accurate IC_{50} curve fitting capacity.

ACKNOWLEDGMENTS

The authors would like to thank Don McKenzie for his support on development and application of AIR Binder. The authors are also grateful to feedback from our clients and friends across pharmaceutical and biotech companies.

REFERENCES

1. Art Carpenter, Carpenter's Complete Guide to the SAS Macro Language, 2nd Edition, SAS Publishing, 2004.
2. Paul R. Ortiz de Montellano, Cytochrome P450: Structure, Mechanism, and Biochemistry, 3rd Edition, Springer, 2004.
3. R. Scott Obach, et al., Mechanism-Based Inactivation of Human Cytochrome P450 Enzymes and the Prediction of Drug-Drug Interactions, Drug Metabolism and Disposition, 35: 246-255, 2007.

ABBREVIATIONS

ADME, Absorption, Distribution, Metabolism and Excretion; CRO, Contract Research Organization; DDI, Drug-Drug Interaction; IND, Investigational New Drug; NDA, New Drug Application; ODS, Output Delivery System; QC, Quality Control; SOP, Standard Operating Procedure

CONTACT INFORMATION

Hao Sun, PhD
Drug Metabolism and Biotransformation
Covance, Inc.
3301 Kinsman Blvd, Madison, Wisconsin 53704
Email: hao.sun@covance.com
Phone: 608.241.7294