

表1 Cremer 毒性クラス 1,2 で NOAEL 比 5 倍以上となった 28 日間試験(TG407)実施物質の毒性プロファイル

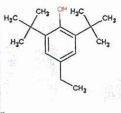

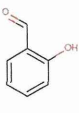

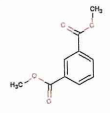
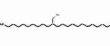
CAS RN	Chemical Name	NOAEL_28d (mg/kg/day) (dose level of main study)	Critical endpoint at LOAEL of main study	NOAEL_14d (mg/kg/day) (dose level of dose setting study)	Critical endpoint at LOAEL of dose setting study	NOAEL ratio (NOAEL_14d/ NOAEL_28d)	Chemical structure	Toxicity class of Cramer rules
4130-42-1	2,6-Bis(1,1-dimethylethyl)-4-ethylphenol	15 (15, 60, 250)	Liver: Hypertrophy, Thyroid: Hypertrophy, follicular cell(60m,f)	125 (125, 250, 500, 1000)	Liver, Kidney weight↑	8.3		Intermediate (Class II)
4286-23-1	4-(1-Methylethenyl)phenol	30 (30, 100, 300, 400)	Fore stomach: Hypertrophy of squamous cells	300 (30, 100, 300)	-(LD50=600)	10.0		Low (Class I)

表2 Cremer 毒性クラス 1,2 で NOAEL 比 5 倍以上となった反復投与生殖発生併合試験(TG422)実施物質の毒性プロファイル

CAS RN	Chemical Name	NOAEL_28d (mg/kg/day) (dose level of main study)	Critical endpoint at LOAEL of main study	NOAEL_14d (mg/kg/day) (dose level of dose setting study)	Critical endpoint at LOAEL of dose setting study	NOAEL ratio (NOAEL_14d/ NOAEL_28d)	Chemical structure	Toxicity class of Cramer rules
90-02-8	2-Hydroxybenzaldehyde	10 (2.5, 10, 40, 160)	Liver: lipid droplet↓ (m), glycogen deposit↑(f) at 40mg/kg	100 (50, 100, 200, 400)	death(200(m1/5,f2/5), 400(m5/5,f4/5))	10.0		Low (Class I)
108-87-2	Methylcyclohexane	62.5 (62.5, 250, 1000)	Cho ↑ (62.5m), Liver weight (1000m,f)	500 (200, 500, 1000)	Liver weight, TP, Cho↑(1000m,f)	8.0		Low (Class I)
1459-93-4	1,3-Benzenedicarboxylic acid, dimethyl ester	62.5 (62.5, 250, 1000)	Weak suppression of body weight increase was seen on female of reproductive test group at more than 250 mg/kg. But no effect on body weight of female of satellite group.	1000 (500, 1000)	-	16.0		Low (Class I)
58670-89-6	2-Decyl-1-tetradecanol	62.5 (62.5, 250, 1000)	PT ↑, LDH ↑ (250m), UrinCl, K, Na ↓ (1000f)	1000 (100, 300, 1000)	-	16.0		Low (Class I)

平成 27 年度 厚生労働科学研究費補助金（化学物質リスク研究事業）  
分担研究報告書

研究課題名：化学物質のヒト健康リスク評価における（定量的）構造活性相関および、  
カテゴリーアプローチの実用化に関する研究

分担研究課題名：反復投与毒性を指標にした構造活性相関モデルに関する研究

研究分担者	広瀬 明彦	国立医薬品食品衛生研究所	安全性予測評価部	部長
研究分担者	小野 敦	国立医薬品食品衛生研究所	安全性予測評価部	第一室長
研究協力者	山田 隆志	国立医薬品食品衛生研究所	安全性予測評価部	第四室長
研究協力者	高橋 美加	国立医薬品食品衛生研究所	安全性予測評価部	研究員
研究協力者	松本 真理子	国立医薬品食品衛生研究所	安全性予測評価部	研究員
研究協力者	川村 智子	国立医薬品食品衛生研究所	安全性予測評価部	研究員
研究協力者	Alex Cayley	ラーサ研究所		
研究協力者	Lilia Fisk	ラーサ研究所		
研究協力者	Sebastien Guesne	ラーサ研究所		
研究協力者	Rachael Tennant	ラーサ研究所		
研究協力者	Jonathan Vessey	ラーサ研究所		

## 要旨

H27 年度は、これまでの公表論文等のメカニズム情報を基に肝毒性アラートを作成する手法に代わる新たなアプローチとして、肝毒性エンドポイントに関連する **key event** モデルを予測し、生体内経路の攪乱に關与する化合物群のプロファイルと組み合わせることによる肝毒性の予測モデルの作成を行った。まず、肝毒性に関連する経路および **Key Event** を特定するために、*in vitro* 肝毒性データを収集した MIP-DILI プロジェクトの研究成果を基礎として MIE (**molecular initiating event**) を特定し、有害転帰経路 (AOP) に位置付けた。さらに、AOP wiki に記載された情報を用いて **key event** と AOP との相関関係を構築した。概念実証プロファイラーの作成には 2 種類のアプローチ (Derek Nexus のアラートパターンの利用と **key event** に関連する公開 *in vitro* アッセイデータに関連する化合物群との構造的類似性を利用したリードアクロス法) を採用した。次に **key event** モデルと化合物プロファイラーを関連づけ、KNIME 解析プラットフォームを用いて肝毒性の予測性能を評価した。反復投与毒性データセットをこのプロファイラーで処理して毒性予測を行った結果、感度は 57% で、特異度は 44% となった。さらに、高濃度でしか肝毒性を示さなかった化合物として LOAEL が 1000 mg/kg 超または 250 m/kg 超となる化合物を除外するとモデルの感度はそれぞれ 63% および 71% に向上することが示された。さらに、物理化学的な記述子のカットオフとして分子量の 480 未満または ClogP-5 超を利用したところ、感度を同程度に維持したまま特異度が 44% から 58% に向上することができた。今後の研究では、感度と特異度の両方を向上させるため、偽陰性や偽陽性の化学物質クラスの解

析等を行うことによって、様々な key event を予測する方法をさらに調査することが有用であることが示された。

## A. 目的

本研究に先立つ研究において、げっ歯類反復投与毒性データセットを用いて作成した肝毒性のラピッドタイプアラートをフルアラートへと発展させる研究を行ってきており、Derek Nexsus における検出感度を、7%から 40%に上昇させることに成功してきた。しかし、これまでの、病理組織学的データや公知のデータセット、公表論文の調査結果を用いて一つ一つアラートを検討するために相当の時間がかかることが問題点として挙げられてきた。そこで、今年度は、反復投与毒性の予測に対する新たなアプローチを確立し、これに基づいて肝毒性に関する概念実証モデルを作成することを試みることにした。まず、複数のモデルを作成して肝毒性エンドポイントに関連する生体内経路において攪乱される key event を予測することとした。さらにこれらのモデルを生体内経路の攪乱に関与する化合物群のプロファイルと組み合わせることにより、対象化合物の肝毒性の予測に活用することができると考えた。

## B. 方法

### B-1 肝毒性に関連する経路および Key Event の特定：

肝毒性に関連する key event は、Lhasa 社によるプロジェクト (MIP-DILI プロジェクト) の一環として行われた研究を基礎として特定した。データは ChEMBL、PubChem などの様々な公開データソースから集められと共に Tox21 のデータセット (<http://www.ncats.nih.gov/tox21>) や他の公開データセット (Aleo et al., Thompson et al., Sakatis et al., Xu et al., Pedersen et al., Dawson et al., Morgan et al.) から得られたデータも含めた。次にこれらの key event および key event に

関して収集したデータを有害転帰経路 (AOP) に位置付けた。さらに、AOP wiki ([https://aopkb.org/aopwiki/index.php/Main\\_Page](https://aopkb.org/aopwiki/index.php/Main_Page)) および公開文献 (Vinken et al., Mellor et al., Jennings et al.) に記載された AOP に関する情報を用いて、key event と AOP との相関関係を構築した。これによりその攪乱が肝毒性の原因になると考えられる経路のネットワークを特定し、経路に関する知識および key event と経路との相関関係を Cytoscape で記録した。(図 1)。

### B-2 特定された経路における Key Event のモデル化：

- DX の構造アラートに収録された知識を用いたモデル作成

Key event の攪乱に関して DX 知識ベースに存在する広範な既存知識を利用するため、種々のエンドポイントに関するアラートを分析した。反応性に基づく情報が最も豊富と考えられた変異原性と皮膚感作のエンドポイントについて分析を行い、反応性に基づく新たな MIE (molecular initiating event) のエンドポイントに結び付けた。代謝活性化の必要性についても考慮し、アラートパターンの中に直接反応する化合物と代謝を必要とする化合物が混在していた場合には別々のアラートに分けた。また複数の機序により同一の毒性に至る化学クラスに関するアラートは、各 key event に関連づけた。さらに、他の関連 key event に関する変異原性アラートおよび皮膚感作アラートについても検討し、該当する key event エンドポイント (例：酸化剤、核酸塩基模倣作用など) を付与した。これらのアラートに関する分析に基づき、肝毒性 AOP につながると考えられ

るさらなる key event の相互作用を特定し、可能な場合にはアラートに付与した。さらに、ミトコンドリア機能障害などの key event を既に予測できる Derek のエンドポイントもモデルに含めた。

- 公開文献から得たデータを用いたリードア  
クロスモデルの作成

Derek の構造アラートにコード化された知識を用いた予測に加え、クエリー化合物とこれら key event に関連するアッセイから収集した公開データとの構造的類似性を利用したリードアクロス法を key event の予測に採用した。この作業は KNIME 解析プラットフォームのフレームワーク内で実施した。

ある key event データセットに関連する化合物 1 つ以上に対して 0.7 以上の類似性を示したクエリー化合物について、当該 key event のフラグを付した。

### B-3 肝毒性のエンドポイントをプロファイリングする Key Event モデルネットワークの作成

前述の 2 種類の方法で反復投与毒性データセットの key event プロファイリングを行った後、これらの手法で得た各化合物の結果を統合した。異なる手法で記述した key event の多くは同一のものであったが、結果においては別々のプロファイルとした。リードアクロス法で key event 予測を行ったものには語尾に「RA」を付した。

### B-4 肝毒性の総合的予測を目的とした Key Event プロファイラーの結果の判定

反復投与データセットにおける化合物の key event プロファイリングで得た結果を統合し、各化合物の肝毒性を総合的に予測した。判定は保守的に行い、プロファイラーのいずれか一つの

key event に関連していれば総合的予測は「陽性」とした。いずれの key event にも関連のない化合物は肝毒性に関する総合的予測を「陰性」とした。

## C. 結果

### C-1 肝毒性に関連する経路および Key Event の特定。

前述の方法によって肝毒性のエンドポイントに関連すると現在特定されている経路および key event の概略は図 1 に示した。この図は経路可視化ツール Cytoscape のスクリーンショットであり、影響を及ぼすレベル（分子～組織）によって全ての key event を色分けして示している。Key event は証拠と結び付けられており、その矢印の太さは証拠の強さによって変えている。

このネットワークは、化学物質が「活性種」として肝臓に到達すると攪乱される可能性のある key event を示すものである。現時点では、化合物が標的臓器に到達するために必要な段階や標的臓器と相互作用を起こしたり不活化されたりするために必要な代謝などは考慮されていない。key event に関して構築したモデルおよびこのモデルを用いた肝毒性のエンドポイントの推定においては吸収、分布、代謝、排泄を一部考慮する必要がありと考えられ、今後の研究で本ネットワークにも収録すべきである。また、このネットワークで用いた概念大系や用語、さらに key event を記述したレベルについてもまだ議論の余地があり、今後検討する必要がある。

### C-2 特定された経路における Key Event のモデル化

- DX の構造アラートに収録された知識を用いたモデル作成

前述の key event の情報を確認するため既存の DX 知識ベースを解析し、肝毒性に関連したアラートと関連する key event のリストを作成した。

DX 知識ベースで確認された key event および各 key event に関連するアラートの数を表 1 に示す。なお key event の攪乱に代謝活性化を要する化合物クラスには (M) を付した。

割り当てられたアラートの数が最も多かった key event は反応型の MIE であるタンパク反応性求電子剤 (M)、DNA 反応性求電子剤 (M)、タンパク反応性求電子剤、DNA 反応性求電子剤であり、その多くは DX における変異原性アラートや皮膚感作アラートから得られた。またこれらのエンドポイントは酸化剤および酸化剤 (M) といった key event エンドポイントの主な原因でもある。ミトコンドリアへの影響は、「ミトコンドリア機能障害」のエンドポイントに対するアラートに加えて「アセチル CoA 欠乏」、「β 酸化阻害」といったより特異的な key event のアラートによっても示されている。

#### ● 公開文献から得たデータを用いたリードア クロスモデルの作成

前述の方法で概説した基準を用いて構築したネットワーク内の key event に関連する公開データを確認し、30 の key event に関するデータを収集した。データが確認された key event を表 2 に示す。この表より、関連するアッセイで試験された化合物が比較的多い key event もあれば、ほとんどない key event もあることが分かる。例えばエストロゲン受容体 (ER) 結合に関するデータは 2233 化合物で存在するが、細胞毒性に関するデータは 6 化合物でしか存在しない。このことは明らかに各リードアクロスモデルがカバーする化学スペースに依存するが、予測対象化合物がどれか一つの構造に類似していればリードアクロスを行うことは可能であると考えられる。表 2 は後回情報の著者の基準に従い陽性と判断された化合物の数を示しており、リードアクロスにはこれらの化合物のみを用いた。大抵の場合、公開文献では陰性と判断されたり、基準を満たさなかった

りする化合物が多くあったが、これらの化合物はリードアクロスや活性の判定には使用しなかった。また、リードアクロスからの陰性データポイントの除外は、クエリー化合物と類似する陰性化合物が陽性化合物と同じクラス内にある場合には当該クエリー化合物が key event をアクティブ化する可能性は否定できないという原則に基づいて行った。

#### ● 肝毒性 Key Event プロファイラーによる反 復投与毒性データセットの解析

B-3 で示した方法で作成した key event モデルを用いて、反復投与毒性データセットにおける化合物の肝毒性を予測した。一つの key event に対してある予測対象化合物が適合した場合には、予測方法は問わず、当該化合物を陽性と予測するのに十分であるとみなした。この予測方法は非常に保守的であると考えられる。予測を行った結果を表 3 に示すが、高水準の予測性が得られた。これらの結果は、データセットに対する妥当な感度 (57%) および予測に関する明確な根拠説明が維持されていることを示している。一方、予測の特異性は低かった (44%)。予測に用いたモデルは ADME 特性を考慮しておらず、この ADME 特性により key event プロファイリングの過程で特定された key event の攪乱が無効化される可能性があるため、特異度の低さはある程度予期された。ADME を考慮に入れたポストフィルタリング等を行うことにより特異度性が向上する可能性がある。

感度は作成したモデルが既知の key event のどのくらいをカバーし、これら key event に関して得られた知識がどのくらい包括的であるかの指標となる。すなわち 57% という感度は肝毒性に至る key event に関して改善の余地があることを示しており、これを達成するため偽陰性を解析し当該化合物が key event プロファイラーで予測されなかった理由を特定することが必要である。

一方、肝毒性を有する可能性のある化合物の多

くが、特異的な経路ではなく高濃度での曝露によって肝臓に広範な負担を与える結果毒性を発現する可能性に着目した。このケースが今回の解析研究のデータに当てはまるか確認するため、高濃度でしか肝毒性を示さなかった陽性化合物 (LOAEL が 1000 mg/kg 超または 250 mg/kg 超) を除外して、予測モデルの感度にどのような影響が及ぶかも調査した。表 3 に示すように、LOAEL 1000 mg/kg 超の化合物をカットオフとするとモデルの感度が向上し (63%) し、カットオフの基準を 250 mg/kg 超まで下げるとさらに感度が上がった (71%)。カットオフを 250 mg/kg まで下げた場合に予測から除外された化合物の大部分は、偽陰性 (FN) の化合物であった。

反復投与肝毒性データセットに関して種々の key event モデルで陽性と予測されたものを図 2 および図 3 に示す。図 2 は DX アラートから予測された key event、図 3 はリードアクロスで予測された key event を示している。これらの図より、モデルが予測した key event の大部分が反応性に基づく key event (タンパク反応性求電子剤、DNA 反応性求電子剤など) が占めていた。酸化剤の key event も相当数の化合物に関して予測されている。一方、化合物のプロファイリングへのリードアクロスの寄与はずっと小さい (Figure 3)。またこれらの key event モデルが真陽性化合物の予測数と同程度の偽陽性化合物数も予測していた。各 key event モデルの陽性率の高さから、核内受容体結合に関連するモデルの多く、およびグルタチオン枯渇や付加物形成に関連する key event モデルは肝毒性のエンドポイントを良好に予測することが示された。一方、求電子反応に関連する key event モデルの陽性適中率は低く、これは同 key event モデルが ADME 特性を考慮していないことが原因と考えられた。

#### ● 偽陰性の調査

偽陰性 (FN) の解析では、構造解析ツール

(Mercury) およびヒトの視覚的分析によるクラスタリングを行い、1) クラス内に 3 種以上の化合物、2) 陽性適中率 50%超のものを選別した結果、18 の化合物クラス (合計 124 物質) が選択された。各化学クラスに関連する肝毒性に関する毒性データおよび機序を探索した。各化学クラスに関する毒性データは、ChemIDPlus、HESS、Google などの公開されているデータベースで検索した。その結果、12 のクラスには毒性と関連するデータを得ることができなかったが、4 つのクラスは、既存の key event (エストロゲン受容体、酸化剤、反応性求電子剤、β 酸化阻害など) に関連する化合物と構造が類似しており、これまでの化学構造の範囲の GAP を補える可能性が示された。また一つクラスは (サリチル酸またはアナロと類似) は新たな MIE を開発できる可能性を示した。残りの一つは動物実験による肝毒性との関連性は低いことが想定された。これらの解析は、予備的な段階であり、今後より詳細な情報収集を行って、肝毒性と関連するアラート構築に向けた研究が必要である化学物質リストとなるものと考えられた。

#### ● 偽陽性の調査

前述のように、反復投与毒性データセットに対する肝毒性 key event プロファイラーの感度は良好であった一方で本モデルの特異度 (44%) には改善の余地がある。これらは、ADME 特性や、代謝による肝毒性の無毒化の可能性を考慮していないことに起因すると考えられた。そこで、ADME 特性を表現することのできそうな包括的な記述子を予備的に調査した。プロファイラーの一つ以上の key event に関連づけられていることを根拠として肝毒性を有すると予測された全ての化合物に関し、分子量、ClogP、極性表面積 (TPSA) および化合物が Lipinski's rule of 5 から外れた回数を計算した。次に化合物を真陽性 (実験に基づき陽性) と偽陽性 (実験に基づき陰性) に分け、前述の記述子についてこれら 2 つの

群の区別に用いることのできるカットオフを見出すための解析を行った。その結果、ほとんどの記述子で実験に基づき陰性である化合物を除外し陽性化合物の除外は最小限とするカットオフを選択することができた。ADME 特性の観点からは非常に大きい化合物は容易に吸収されないであろうと考えられると同様に、高親水性の化合物も吸収されるとは考えにくい。記述子ごとに異なるカットオフが真陽性と偽陽性の区別に及ぼす影響を表 4 に示した。真陽性をあまり除外することなく大部分の偽陽性を除外するという点では、分子量のカットオフが最善の区別方法であると考えられた。さらに分子量と ClogP の最小値を組み合わせるとより多くの偽陽性を除外することができた。これらの記述子によるカットオフをポストフィルタとして組み入れた後のモデルの予測性能は表 5 に示した。この表より、いずれの場合にも感度の低下は最小限で特異度が上昇し、結果として **balanced accuracy** も上昇した。

#### D. 考察

今回の解析で化合物が肝毒性に関連する経路の **key event** を攪乱する可能性を予測することは、当該化合物が肝毒性を誘発する可能性をプロファイリングする方法として有望であることが示された。このネットワークを用いて反復投与毒性データセットにおける肝毒性を有する化合物の大部分 (57%) を確認することができ、さらにこの特定により、攪乱する **key event** ごとに当該化合物が活性を示すと考えられた根拠に関する知見も得ることができた。一方、プロファイラーによる肝毒性予測の特異度は比較的低かったが、これは肝毒性に至ると仮定した相互作用を妨げる可能性のある因子 (例、ADME 特性) を考慮していなかったことによるものであることは想定された。この特異度が少なくとも活性に関する基礎的な ADME の要件をモデル化した単純な記述子とカットオフを用いることで改善できることは、既に

予備的調査で明らかにした。また、偽陰性を分析した結果、ネットワークで予測されない NIHS データセットにおける化学クラスの多くが未知の機序により肝毒性を発現する可能性が示された。これは、肝障害を引き起こすものとして本研究で採用した MIE が主として医薬品に関して調査されているものを基にしていることが原因と考えられた。今後の研究では、予測の正確度および適中率を向上 (従って、感度と特異度の両方を向上) させるため、様々な **key event** を予測する方法をさらに調査することが有用であろう。また、プロファイラーが「正しい」理由によって正しい予測を導いているかをさらに検討 (すなわち、化合物の肝毒性に関して妥当な根拠を確認) し、さらに潜在的な危険有害性が肝毒性のリスクに変換されるのかを予測するため化合物の物理化学的性質を用いてモデルの特異度を向上させることも有益であると考えられる。最後に、現行のシステムでは未だモデル化されていない肝毒性に関連する **key event** および経路を特定、モデル化することによってモデルの感度を向上させることも可能であると考えられた。

#### E. 結論

H27 年度は、これまでの公表論文等のメカニズム情報を基に肝毒性アラートを作成する手法に代わる新たなアプローチとして、肝毒性エンドポイントに関連する **key event** モデルを予測し、生体内経路の攪乱に関与する化合物群のプロファイルと組み合わせることによる肝毒性の予測モデルの作成を行った。予測モデル作成のための解析には KNIME プラットフォームを用い化合物プロファイラーを作成した。反復投与毒性データセットをこのプロファイラーで処理して毒性予測を行った結果、感度は 57% で、特異度は 44% となった。さらに、高濃度でしか肝毒性を示さなかった化合物を除外するとモデルの感度は 71% まで向上することが示された。さらに、分子量の 480 未満または ClogP-5 超という物理化学的記述子の

カットオフを適用したところ、感度を同程度に維持したまま特異度を 58%に向上することができた。今後は、偽陰性や偽陽性の化学物質クラスの解析等を行うことによって、様々な key event を予測する方法をさらに調査することが有用であることが示された。

## F. 研究発表

### 1. 論文発表

Hirata-Koizumi, M., Fujii, S., Kato, H., Matsumoto, M., Takahashi, M., Ono, A. and Hirose, A., Repeated dose and reproductive/developmental toxicity of long-chain perfluoroalkyl carboxylic acids in rats: perfluorohexadecanoic acid and perfluorotetradecanoic acid. *Fundam. Toxicol. Sci.*, 2, 177-190, 2015.

Ono, A., Kobayashi, K., Serizawa, H., Kawamura, T., Kato, H., Matsumoto, M., Takahashi, M., Hirata-Koizumi, M., Matsushima, Y. and Hirose, A., A repeated dose 28-day oral toxicity study of  $\beta$ -bromostyrene in rats *Fundam. Toxicol. Sci.*, 2, 191-200, 2015.

Canipa S, Cayley A, Drewe WC, Williams RV, Hamada S, Hirose A, Honma M, Morita T. Using in vitro structural alerts for chromosome damage to predict in vivo activity and direct future testing. *Mutagenesis*. 2016 31: 17-25.

高橋美加、松本真理子、宮地繁樹、菅谷芳雄、長谷川隆一、小林克己、平田睦子、小野敦、広瀬明彦. OECD 化学物質対策の動向 (第 26 報) - 第 6 回 OECD 化学物質共同評価会議 (2014 年パリ) 化学生物総合管理, 11, 28-36, 2015

松本真理子、清水将史、宮地繁樹、菅谷芳雄、広瀬明彦. OECD 化学物質共同評価プログラム: 第 6 回化学物質共同評価会議概要 化

学生物総合管理, 11, 37-45, 2015.

### 2. 学会発表

Hirose A. Japanese Current Chemical Regulation and Contribution to the OECD Cooperative Chemicals Assessment Programme (CoCAP). The 7th International Congress of Asian Society of Toxicology. (2015.7. チェジュ、韓国)

Hirose A, Hirata-Koizumi M, Kawamura T, Matsumoto M, Takahashi M, Nishimaki-Mogami T, Nishimura T, Ema M, Ono A. Derivation of subacute reference doses for drinking water quality management. The 51st Congress of the European Societies of Toxicology (EUROTOX 2015) (2015.9. ポルト, ポルトガル)

Ono, A., Matsumoto, M., Takahashi, M., Kawamura, T., Hirata-Koizumi, M. and Hirose, A. 2015. Is a 14-day dose setting study able to predict its 28-day repeated dose toxicity? The 51st Congress of the European Societies of Toxicology (EUROTOX 2015) (2015.9. ポルト, ポルトガル)

Hirose A. Metal contaminants in drugs: ICH point of view. The 9th Congress of Toxicology in Developing Countries. (2015.10. ナタール, ブラジル)

Nishimura T, Hirata-Koizumi M, Yamada T, Kawamura T, Ono A, Hirose A, Ema M. Derivation of the health advisory guidance values for sub-acute exposure of drinking water. The Society of Toxicology 55th Annual Meeting (SOT 2016) (2016.3. ニューオーリンズ)

## G. 知的財産権の出願・登録状況



1. 特許取得: 該当なし
2. 実用新案登録: 該当なし
3. その他: 該当なし

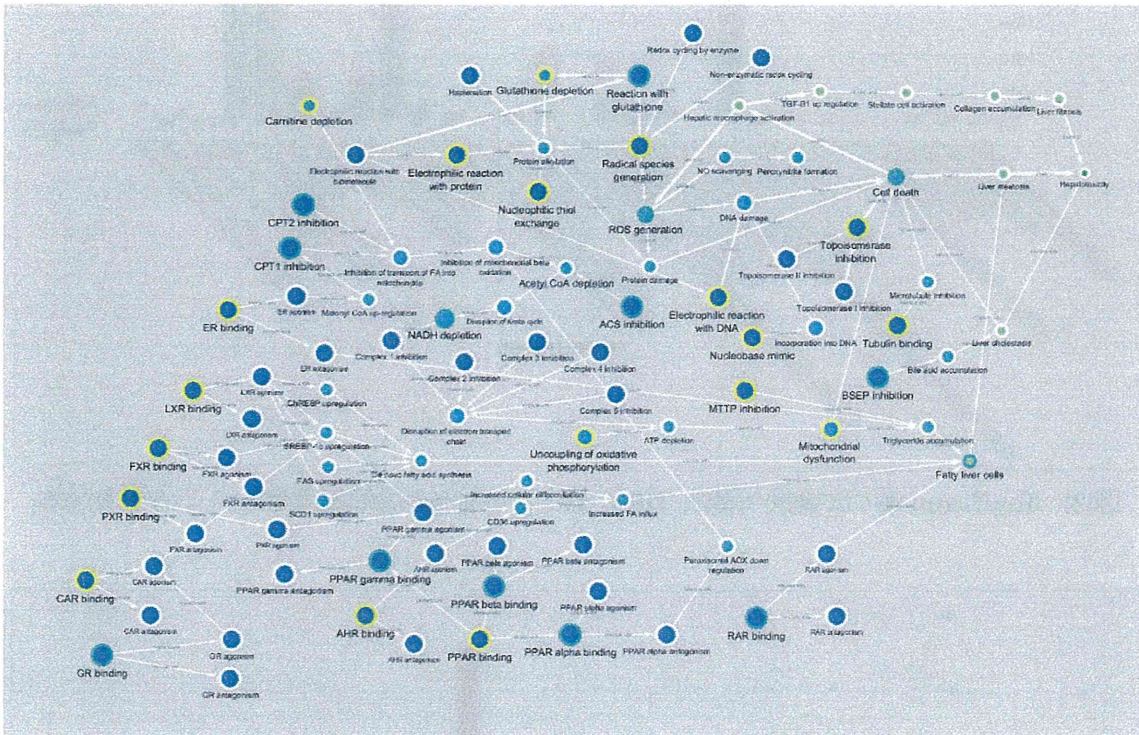


図 1. 経路可視化ツール Cytoscape でモデル化した肝毒性の発現に関連する key event ネットワークのスクリーンショット。

各点は分子レベル、細胞小器官レベル、細胞レベル、組織レベル、臓器レベルで色分けされている。点の周囲が黄色でハイライトされている key event は DX の構造アラート、濃緑色のハイライトはリードアクロス (RA)、黄緑色のハイライトは DX と RA の両方を用いてモデル化された。

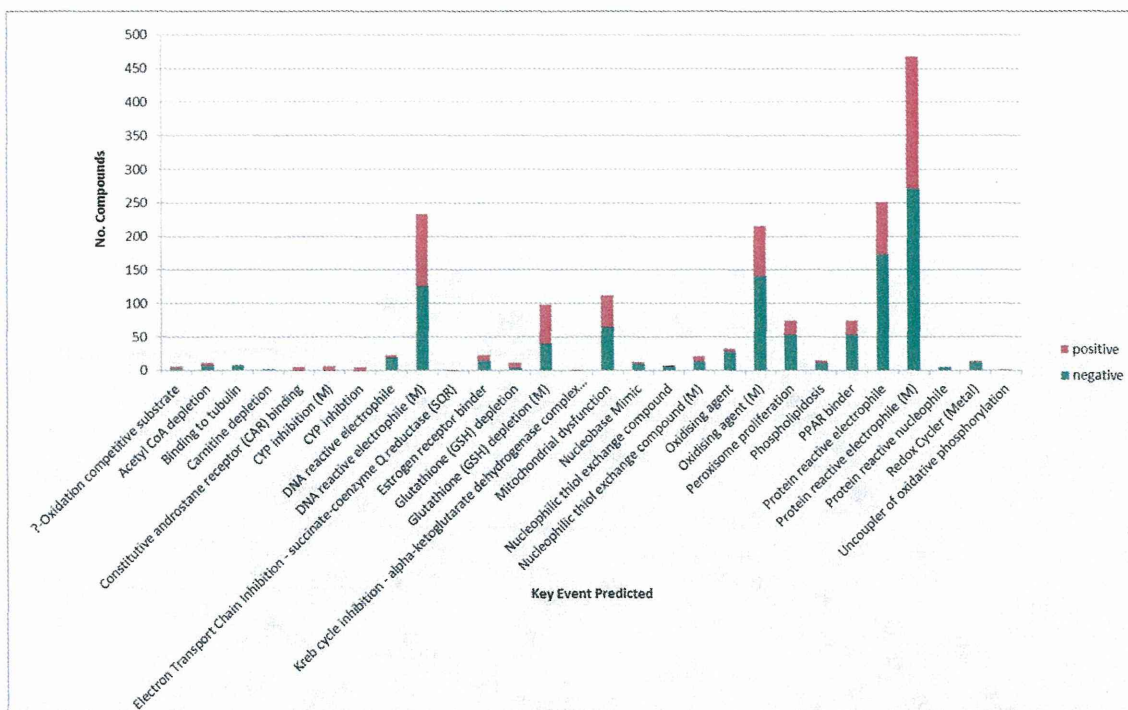


図 2. KeyEvent 毎の反復投与毒性 DB における肝毒性の陽性物質数および陰性物質数

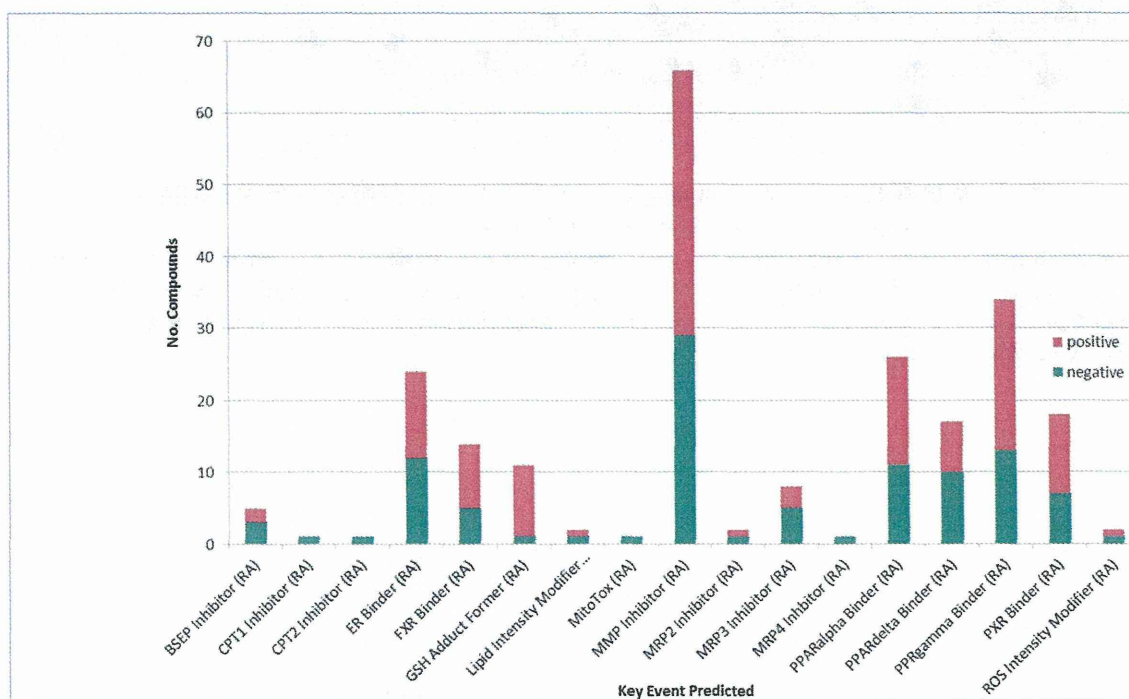


図 3. リードアクロス解析における KeyEvent 毎の反復投与毒性 DB における肝毒性の陽性物質数および陰性物質数

表 1. 肝毒性に関連づけられた DX に記載されている Key Events

Key Event	No. Alerts Predicting Event
Protein reactive electrophile (M)	61
DNA reactive electrophile (M)	58
Protein reactive electrophile	51
DNA reactive electrophile	44
Mitochondrial dysfunction	24
Oxidising agent	20
Oxidising agent (M)	20
Estrogen receptor binder	17
Glutathione (GSH) depletion (M)	14
PPAR binder	7
Peroxisome proliferation	7
Phospholipidosis	5
Nucleobase Mimic	4
Uncoupler of oxidative phosphorylation	4
Acetyl CoA depletion	3
Aryl hydrocarbon receptor binder	2
CYP inhibition (M)	2
Electron Transport Chain Inhibition	2
Microsomal Triglyceride Transfer Protein (MTTP) antagonist	2
Nucleophilic thiol exchange compound (M)	2
$\beta$ -Oxidation competitive substrate	2
$\beta$ -Oxidation inhibition	2
3-Hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibition	1
Angiotensin converting enzyme (ACE) inhibitor	1
Bile transport competitive substrate (M)	1
Binding to tubulin	1
CYP inhibition	1
Carnitine depletion	1
Carnitine palmitoyltransferase I inhibition	1
Chelation of metal ions	1
Constitutive androstane receptor (CAR) binding	1
Electron Transport Chain Inhibition - succinate-coenzyme Q reductase (SQR)	1
Glutathione (GSH) depletion	1
Inhibition of RNA polymerase (M)	1
Krebs cycle inhibition - alpha-ketoglutarate dehydrogenase complex inhibition	1
Low density lipoprotein (Lys-adduct) reactive electrophile (M)	1
Nucleophilic thiol exchange compound	1
Protein reactive nucleophile	1
RNA reactive electrophile (M)	1
Redox Cyclers (Metal)	1

表 2 公表文献から得られた肝毒性に関連する KeyEvents と  
リードアクロスに用いた陽性化合物数

<b>Key Event</b>	<b>No. of Positive Compounds Reported</b>
ER Binder (RA)	2233
GR Binder (RA)	1853
Tubulin Binder (RA)	705
LXR Binder (RA)	664
Topoisomerase Inhibitor (RA)	582
RXR Binder (RA)	408
MMP Inhibitor (RA)	385
RAR Binder (RA)	235
BSEP Inhibitor (RA)	213
AHR Binder (RA)	197
MTTP Binder (RA)	185
MRP2 Inhibitor (RA)	179
NADH Depletor (RA)	133
PXR Binder (RA)	122
FXR Binder (RA)	98
GSH Adduct Former (RA)	97
MRP4 Inhibitor (RA)	83
PPRgamma Binder (RA)	77
P450 Inhibitor (MDI assay) (RA)	74
MitoTox (RA)	53
CAR Binder (RA)	50
CPT1 Inhibitor (RA)	39
PPARdelta Binder (RA)	36
ROS Intensity Modifier (RA)	32
MRP3 Inhibitor (RA)	30
Lipid Intensity Modifier (RA)	25
PPARalpha Binder (RA)	22
CPT2 Inhibitor (RA)	20
CYP Inhibitor (RA)	10
CytoTox Causing (RA)	6

表 3・ 肝毒性 Key Event プロファイラーによる反復投与毒性データの予測性

陽性 Cut-Off	BA (%)	A (%)	Sen. (%)	Spec. (%)	PPV (%)	NPV (%)	Total	TP	FP	TN	FN
All Data	50	49	57	44	38	63	1458	310	511	402	235
>1000 mg/kg を 除外	54	51	63	44	37	70	1389	301	511	402	175
>250mg/kg を除外	57	52	71	44	35	78	1308	280	511	402	115

表 4. 物理化学的記述子のカットオフが偽陽性の数に及ぼす影響

Descriptor	Cut-off	TP	FP
None	n.a.	310	511
ClogP	-5 to 10	296	437
TPSA	<150	299	416
MW	<580	295	391
Lipinski's rules	<3	302	469
MW, ClogP	MW<580, ClogP>-5	295	387

表 5. 分子量と ClogP 値のカットオフを適用した場合の反復投与毒性データの予測性

陽性 Cut-Off	BA (%)	A (%)	Sen. (%)	Spec. (%)	PPV (%)	NPV (%)	Total	TP	FP	TN	FN
All Data	56	56	54	58	43	68	1458	295	387	526	250
>1000 mg/kg を 除外	59	58	60	58	42	73	1389	286	387	526	190
>250mg/kg を除外	62	61	67	58	41	80	1308	266	387	526	129

III. 研究成果の刊行に関する一覧表

## 研究成果の刊行に関する一覧表

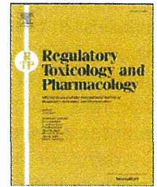
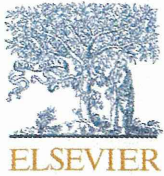
## 雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Barber C, Amberg A, Custer L, Dobo KL, Glowienke S, Van Gompel J, Gutsell S, Harvey J, Honma M, Kenyon MO, Kruhlak N, Muster W, Stavitskaya L, Teasdale A, Vessey J, Wichard J	Establishing best practise in the application of expert review of mutagenicity under ICH M7.	Regul Toxicol Pharmacol.	73	367-377	2015
Petko IP, Patlewicz G, Terry, Schultz W, Honma M, Todorov M, Kotov S, Dimitrov SD, Donner M, Mekenyan OG	A feasibility study: Can information collected to classify for mutagenicity be informative in predicting carcinogenicity?	Regul Toxicol Pharmacol.	72	17-25	2015
Canipa S, Cayley A, Drewe WC, Williams RV, Hamada S, Hirose A, Honma M, Morita T	Using in vitro structural alerts for chromosome damage to predict in vivo activity and direct future testing.	Mutagenesis	31	17-25	2015
Hirata-Koizumi M, Fujii S, Kato H, Matsumoto M, Takahashi M, Ono A, Hirose A	Repeated dose and reproductive/developmental toxicity of long-chain perfluoroalkyl carboxylic acids in rats: perfluorohexadecanoic acid and perfluorotetradecanoic acid.	Fundam. Toxicol. Sci.	2	177-190	2015
Ono A, Kobayashi K, Serizawa H, Kawamura T, Kato H, Matsumoto M, Takahashi M, Hirata-Koizumi M, Matsushima Y, Hirose A	A repeated dose 28-day oral toxicity study of $\beta$ -bromostyrene in rats.	Fundam. Toxicol. Sci.	2	191-200	2015
松本真理子、清水将史、宮地繁樹、菅谷芳雄、広瀬明彦	OECD化学物質共同評価プログラム：第6回化学物質共同評価会議概要	化学生物総合管理	11	37-45	2015
Okamura H, Abe H, Hasegawa-Baba H, Saito K, Sekiya F, Hayashi S, Mirokuji Y, Maruyama S, Ono A, Nakajima M, Degawa M, Ozawa S, Shibutani, Maitani T	The Japan Flavour and Fragrance Materials Association's (JFFMA) safety assessment of acetal food flavouring substances uniquely used in Japan.	Food Addit Contam Part A Chem Anal Control Expo Risk Assess	32	1384-1396	2015



Igarashi Y, Nakatsu N, Yamashita T, Ono A, Ohno Y, Urushidani T, and Yamada H	Open TG-GATEs: a large-scale toxicogenomics database.	Nucleic Acids Res	43(Database issue)	D921-D927	2015
Morita T, Uno Y, Honma M, Kojima H, Hayashi H, Tice RR, Corvi R, Schechtman L	The JaCVAM International Validation Study on the in vivo Comet Assay: Selection of Test Chemicals.	Mutation Research	786-788	14-44	2015
Morita T, Hamada S, Masumura K, Wakata A, Maniwa J, Takasawa H, Yasunaga K, Hashizume T, Honma M	Evaluation of the sensitivity and specificity of in vivo erythrocyte micronucleus and transgenic rodent gene mutation tests to detect rodent carcinogens.	Mutation Research	802	1-29	2016
森田 健	LD50値による毒性評価手法の変遷	中毒研究	28	388-391	2015

#### IV. 研究成果の刊行物・別刷



## Establishing best practise in the application of expert review of mutagenicity under ICH M7



Chris Barber <sup>a,\*</sup>, Alexander Amberg <sup>b</sup>, Laura Custer <sup>c</sup>, Krista L. Dobo <sup>d</sup>, Susanne Glowienke <sup>e</sup>, Jacky Van Gompel <sup>f</sup>, Steve Gutsell <sup>g</sup>, Jim Harvey <sup>h</sup>, Masamitsu Honma <sup>i</sup>, Michelle O. Kenyon <sup>d</sup>, Naomi Kruhlak <sup>j</sup>, Wolfgang Muster <sup>k</sup>, Lidiya Stavitskaya <sup>j</sup>, Andrew Teasdale <sup>l</sup>, Jonathan Vessey <sup>a</sup>, Joerg Wichard <sup>m</sup>

<sup>a</sup> Lhasa Limited, Leeds, UK

<sup>b</sup> Sanofi-Aventis Deutschland GmbH, DSAR Preclinical Safety, Frankfurt, Germany

<sup>c</sup> Bristol-Myers Squibb, Drug Safety Evaluation, New Brunswick, USA

<sup>d</sup> Pfizer, Drug Safety Research and Development, Groton, CT, USA

<sup>e</sup> Novartis Institutes for Biomedical Research, Department of Preclinical Safety, Basel, Switzerland

<sup>f</sup> Janssen, Drug Safety Sciences, Beerse, Belgium

<sup>g</sup> Unilever, Safety and Environmental Assurance Centre, Colworth, Beds, UK

<sup>h</sup> GlaxoSmithkline, Computational Toxicology, Ware, Herts, UK

<sup>i</sup> National Institute of Health Sciences, Tokyo, Japan

<sup>j</sup> FDA Center for Drug Evaluation and Research, Silver Spring, MD, USA

<sup>k</sup> F. Hoffmann-La Roche Ltd., Pharma Research and Early Development, Basel, Switzerland

<sup>l</sup> AstraZeneca, Macclesfield, Cheshire, UK

<sup>m</sup> Bayer, HealthCare, Genetic Toxicology, Berlin, Germany

### ARTICLE INFO

#### Article history:

Received 19 July 2015

Received in revised form

21 July 2015

Accepted 22 July 2015

Available online 4 August 2015

#### Keywords:

ICH M7

Ames

Genotoxicity

Mutagenicity

Expert rule-based

Statistical

*In silico*

### ABSTRACT

The ICH M7 guidelines for the assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals allows for the consideration of *in silico* predictions in place of *in vitro* studies. This represents a significant advance in the acceptance of (Q)SAR models and has resulted from positive interactions between modellers, regulatory agencies and industry with a shared purpose of developing effective processes to minimise risk. This paper discusses key scientific principles that should be applied when evaluating *in silico* predictions with a focus on accuracy and scientific rigour that will support a consistent and practical route to regulatory submission.

© 2015 Elsevier Inc. All rights reserved.

### 1. Introduction

The ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use) is an international group comprised of regulatory authorities and industry across Europe, Japan and the US which has taken a

predominant position in establishing guidelines to support the development and registration of safe and effective medicines. Since 1990, this body has established almost 50 guidelines covering many processes from manufacturing quality to clinical trial design and drug safety. One of these (M7) focusses upon the assessment of potentially DNA-reactive (mutagenic) impurities and is significant in that, it recognises the potential to use *in silico* predictions in lieu of *in vitro* studies. This substitution is not necessarily a simple process because of the inherent uncertainty that exists with any *in silico* prediction. However, the risk of not identifying a potentially DNA-reactive impurity that is subsequently exposed to humans can

\* Corresponding author. Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds LS11 5PS, UK.

E-mail address: [chris.barber@lhasalimited.org](mailto:chris.barber@lhasalimited.org) (C. Barber).

be minimised through the appropriate use of well-constructed models and, crucially, through the use of expert analysis. Whilst this principle is clearly defined in the guidelines, the practical implementation of a process that can ensure consistent, safe and accurate predictions requires some consideration. This paper identifies key questions and approaches that an expert could apply in order to maximise sensitivity in the identification of DNA-reactive materials without necessarily applying an overly cautious approach that reduces specificity and overall accuracy to a point where *in silico* predictions offer little additional value.

The intention of the M7 guideline is to describe how best to identify and control the exposure to (potential) pharmaceutical agent impurities that could cause cancer through direct reaction with DNA. Such impurities may be introduced during the preparation of the active pharmaceutical ingredient (API), during formulation of the final drug product, or through degradation. DNA reactivity can be effectively tested using the *in vitro* bacterial reverse mutation assay (often somewhat loosely referred to as 'the Ames assay'), but this study may be impractical or unnecessary given our knowledge of the endpoint and weight of data already obtained. (Q)SAR is recommended in the ICH M7 guidelines as a high-throughput, state-of-the-art alternative for assessing the mutagenic potential of such impurities. This paper focusses not on the identification of such potential or observed impurities, but on the subsequent (Q)SAR analysis of their likelihood of being DNA-reactive.

The guideline describes the need for two predictive systems; one expert rule-based and the second statistical-based. The application of two systems that use different methods is predicated on the assumption that their predictions will be complementary and that greater sensitivity in detecting potential mutagens will be achieved if they are applied in a manner where a positive prediction from either method leads to a positive conclusion. Applying the models in this manner typically results in a decrease in specificity and overall accuracy; however, some of this decrease can be mitigated through the application of expert knowledge. Indeed, the guidelines make a specific provision for the application of expert knowledge to support or overturn a (Q)SAR prediction, effectively allowing a positive or negative *in silico* conclusion to be challenged through the rational consideration of additional information. While expert assessment has been successfully applied and reported (Dobo et al., 2012; Sutter et al., 2013), the definition of what constitutes expert analysis and how it may be undertaken has until now, been left largely open (Powley, 2015; Greene et al., 2015). Here we attempt to define a framework through which this process may be more clearly understood, although it should be emphasised that we do not anticipate this to be exhaustive or that when applied, experts will necessarily come to the same conclusions.

## 2. Comparison between expert rule-based and statistical systems

The distinction between expert rule-based and statistical systems is not necessarily clear-cut; experts may use statistical models to support their knowledge extraction during the development of an expert rule-based system, and the building of statistical systems undoubtedly benefits from oversight by an expert.

An expert rule-based system is comprised of rules written by humans which allows for the injection of knowledge in addition to known (in)activity of compounds. These include the biological mode(s) of action or expected metabolic transformations together with a chemical understanding of reactivity. This facilitates the provision of a detailed and clearly reasoned explanation for an alert along with references and mechanistic information – all of which can offer significant benefit during expert analysis. The injection of

additional knowledge also allows the model builder to extrapolate from that which is known from the training set, for example to hitherto unseen functionality based upon an understanding of chemical reactivity, or through the incorporation of knowledge from proprietary data that can remain transparent and interpretable without revealing confidential structures. This can help an expert rule-based model maintain strong performance against chemical space more dissimilar to the training set than a statistical system can often achieve. Such an approach can also allow more complex endpoints to be modelled – for example if there are multiple modes of actions within a single dataset, then a simple global statistical model may struggle to correctly learn significant trends.<sup>1</sup> Some more sophisticated expert rule-based systems apply layers of reasoning which enables the model to continue to work in the presence of contradictory information – something which is important when modelling biological data since reproducibility is often not complete.

In contrast, a statistical system must rely upon machine-learning to determine the importance of a descriptor or combination of descriptors subsequently used to predict activity. For it to show a high degree of transparency, some expert knowledge almost always influences the model, most commonly through the choice of descriptors from which the model can learn. This can range from defining a list of descriptors to be used during model building (e.g. a predefined fragment library) or by defining rules by which descriptors should be created from the training set (e.g. fragmentation rules appropriate to the endpoint). Each approach offers benefits and challenges, but the result is still a statistical model provided that the decision as to the final selection of descriptors and impact that they make upon a prediction is learnt by the model through application of the training data. Statistical models tend to work best when predicting compounds of similar structure to those in the training set, and this has given rise to a number of approaches to defining applicability domains. Typically this is a set of criteria implemented by the modeller below which the predicted accuracy is considered insufficient for reliable use. However, this measure should be treated with caution, since even if the approach is explicit and the robustness demonstrated (sadly both are often missing), the modeller cannot know the use to which the model will be put and hence the acceptable level of accuracy that is desired. Models that apply the same descriptors for both predictions and determining the applicability domain and focus upon human interpretable properties that have demonstrable relevance to the endpoint are more likely to be able to provide transparent, relevant and unambiguous estimates of expected accuracy than those statistically tuned to a certain level of performance against a particular test set (Sahigara et al., 2012; Netzeva et al., 2005).

## 3. Ensuring *in silico* models can support expert analysis

Many models of mutagenicity have been developed using a range of methodologies. All have benefited from a large dataset of *in vitro* data tested under well established and mostly consistent conditions. This, together with a clear understanding of the mechanisms of action, has enabled models to make predictions which can be trusted provided the supporting evidence and rationale is readily accessible. Whilst some models may be used to illustrate points, this paper does not endorse any specific system

<sup>1</sup> Some statistical approaches can at least partly address this, for example by clustering compounds before constructing local models and thereby identify significant trends that would be swamped if the dataset was considered as a whole (Hanser et al., 2014).