

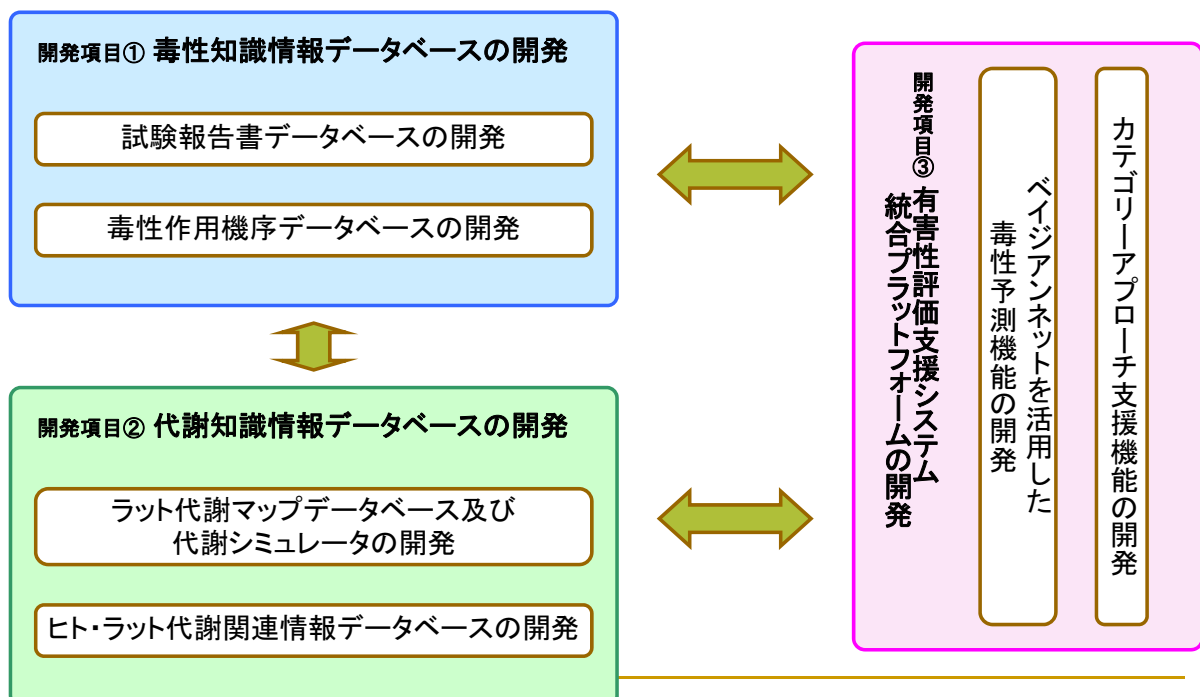
議題5 プロジェクトの詳細説明 (公開)

5.3 有害性評価支援システム 統合プラットフォームの開発

(1) ベイジアンネットワークを活用した 毒性予測機能の開発

2009年8月13日(木)

「構造活性相関手法による有害性評価手法開発」 プロジェクト構成図



1. 開発項目の概要と目標

達成目標

(1) 最終目標 (平成23年度)

■ 毒性知識情報データベース及び代謝知識情報データベースを連携し、毒性予測モデルを組み込み、化学構造を入力すると、代謝物情報や、親物質・代謝物の有害性情報、類似化合物の検索及び解析結果や、肝臓への毒性を中心とした予測結果が表示される有害性評価支援システム統合プラットフォームを開発し、公開する。また、ユーザー独自の情報をデータベースに追加することにより、より精度の向上につながる機能を備えたものとする。開発に当たっては、国際整合性を考慮し、OECD (Q)SAR Application Toolboxとの互換性を確保した仕様とし、カテゴリー化の機能に関しては、OECDで検討されているカテゴリーアプローチの方法論に準拠したものとする。

(2) 中間目標 (平成21年度)

■ 毒性知識情報データベースと代謝知識情報データベースを基に、毒性予測モデルの試作版を開発するとともに、化学構造を入力すると、有害性情報や類似化合物の検索結果が表示され、また、ユーザー独自の情報を追加し、評価する機能も備えたシステム統合プラットフォーム試作版を作成する。

2. 中間目標(平成21年度)に対する達成度

2. 中間目標(平成21年度)に対する達成度

小項目	中間目標(平成21年度末) 出典:基本計画 p.7	研究開発成果	達成度
(2)ペイジアンネットワークを活用した毒性予測機能の開発	毒性知識情報データベースと代謝知識情報データベースを基に、毒性予測モデルの試作版を開発するとともに、化学構造を入力すると、有害性情報や類似化合物の検索結果が表示され、また、ユーザー独自の情報を追加し、評価する機能も備えたシステム統合プラットフォーム試作版を作成する。	試験報告書に含まれる各種検査値を基に、カスケードモデルにより解析を行い、特定の毒性を発現する構造上の特徴を抽出し知識ベースBASiCとして公開した。さらに、芳香族アミン類の反復投与毒性試験における肝毒性及び溶血性貧血を化学構造から推定するためのペイジアンネットワークシステムの試作版を開発した。今後、ネットワークの改良や拡張を行うこと等により、平成21年度中に目標を達成する見込みである。	○

◎:達成、○:達成見込、△:未達、×:中止

3. 進捗状況と成果

目次

- 特徴的活性構造BASの抽出
 - BAS(basic active structure)抽出法の紹介
 - 反復投与毒性を対象とした知識ベースBASiC

- ベイジアンネットによる反復投与毒性評価システム
 - ベイジアンネットとその矛盾解析を利用した予測と評価
 - システム開発の現状と今後の課題

肝毒性データ

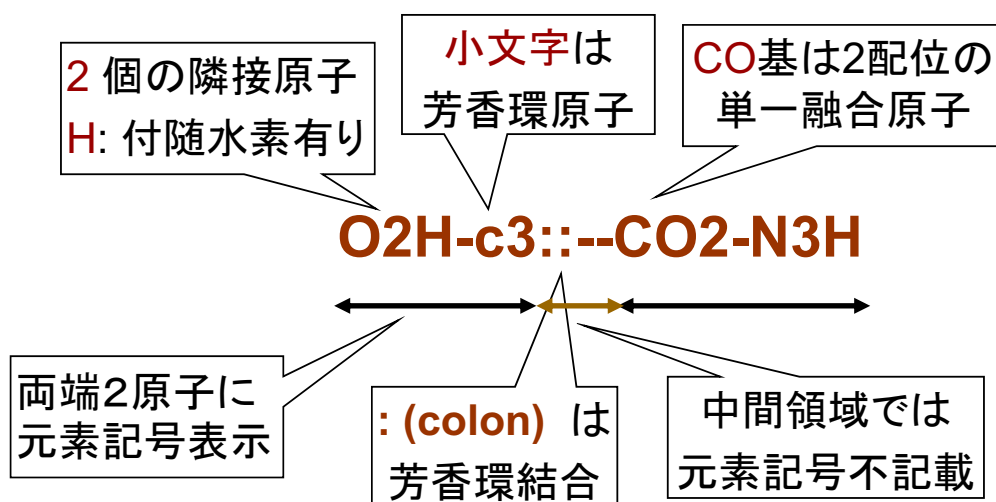
毒性知識情報データベース中の有機化合物141種

GOT	Increase	Normal	Decrease
Male Rat	9	128	4
Female Rat	4	128	9

GPT	Increase	Normal	Decrease
Male Rat	23	114	4
Female Rat	24	113	4

肝相対重量	Increase	Normal	Decrease
Male Rat	72	68	1
Female Rat	84	57	0

線形フラグメントの例

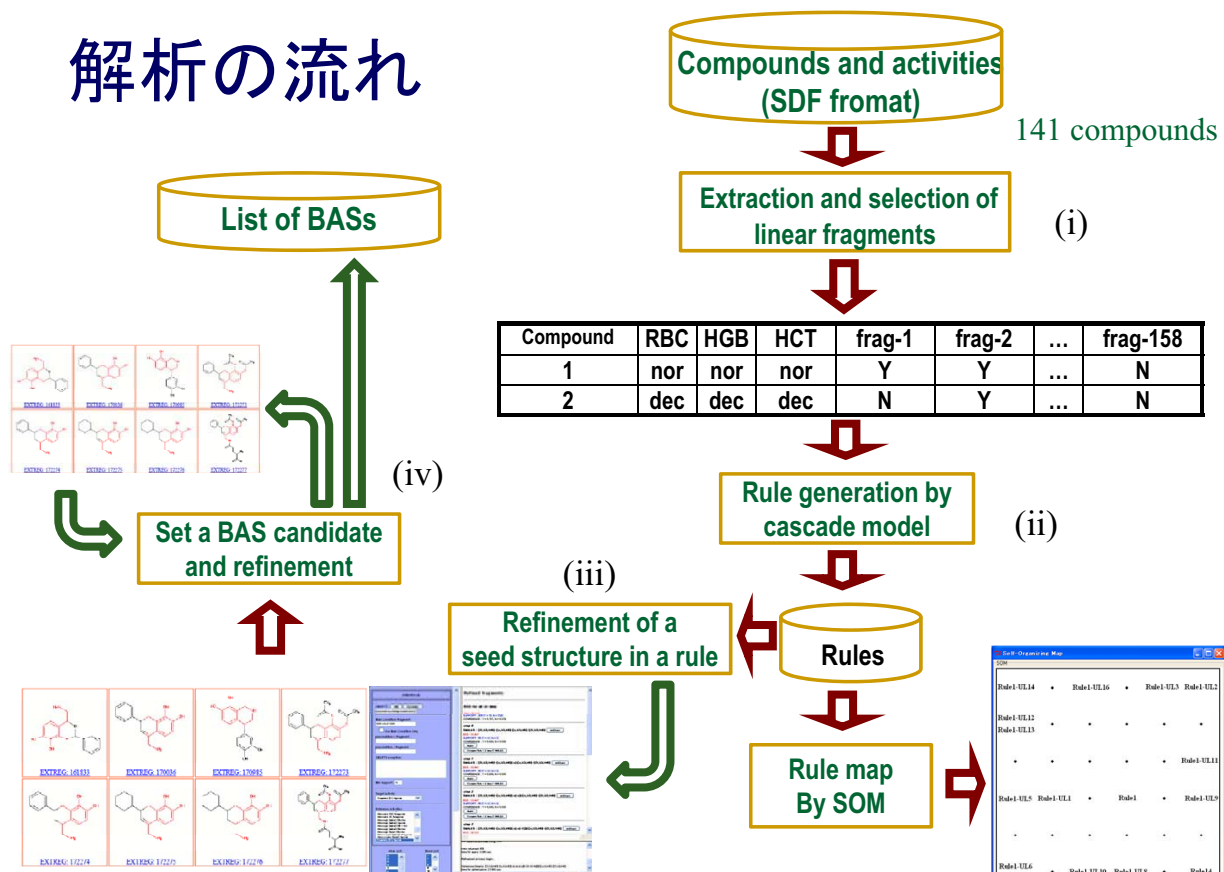


- ◆ SMILES表記にならない化学者が理解容易
- ◆ 詳細過ぎる表記を避ける

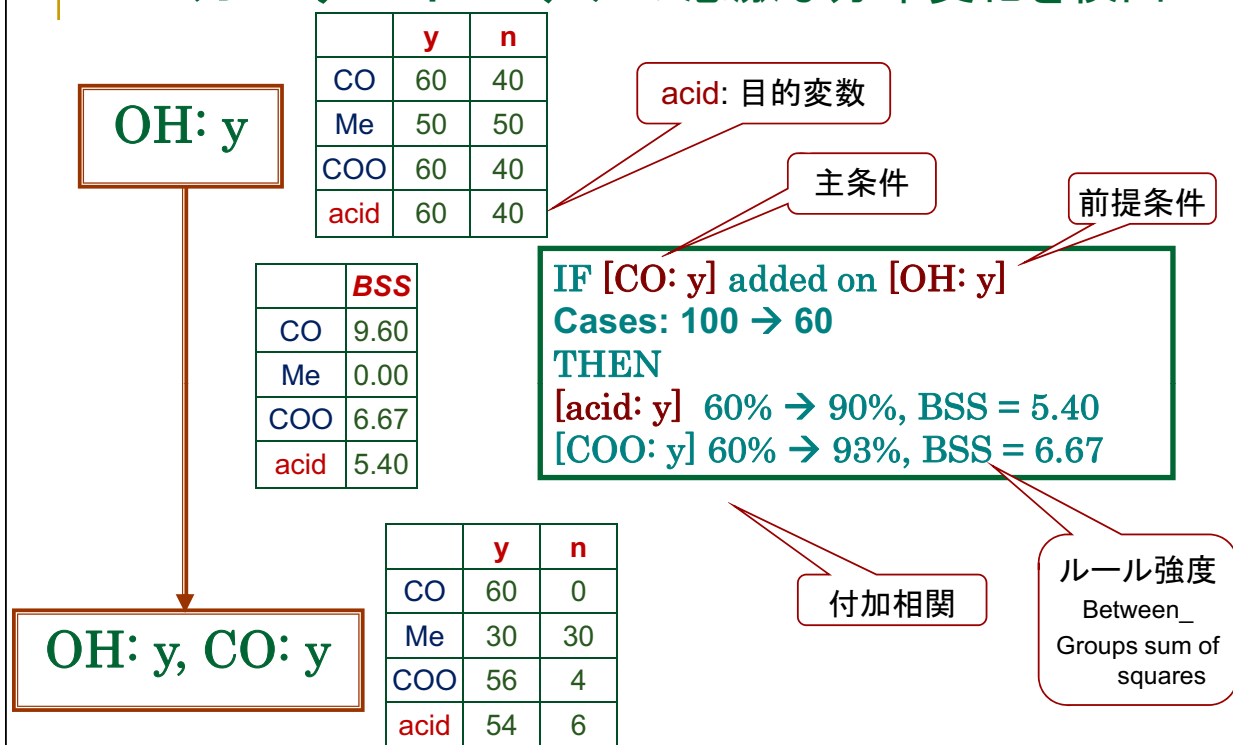
線形フラグメントの選択

1. 生成されたフラグメント：3224種
2. 出現頻度が下記範囲にあるフラグメント259種を選択
 $3\% < P(\text{fragment}) < 97\%$.
3. フラグメント対で出現の相関が大きい ($R_{ij} > 0.9$) 場合一方を採用し, フラグメント150種を選択
4. 相関が高くとも解釈に重要な8種を追加
5. 最終的に158種のフラグメント群を記述子として採用

解析の流れ



カスケードモデル: 急激な分布変化を検出



雌ラットのGOTでもっとも強いルール

Rule 1: **Cases: 60 → 3** トルエン様構造で3化合物

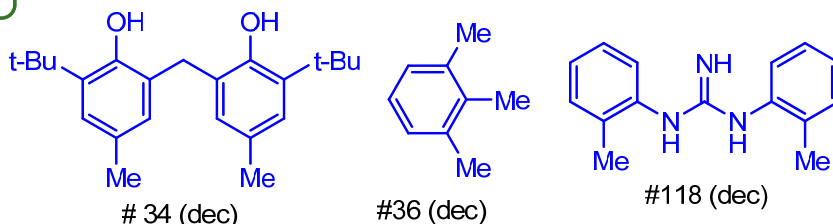
IF [C4H-c3: y]
 added on [C4H-c3::c3H:c3: n] [N3H-c3::c3:c3H: n] [O1: n] [c3H:c3:c3H:c3H: n]
 Then GOT_FR: 0.93 0.00 0.07 ==> 0.00 0.00 **1.00** (none inc dec)

then LIVER_Wt_FR: 0.55 0.45 ==> **1.00** 0.00 (inc none)
 then C4H-c3:c3: 0.05 0.95 ==> **1.00** 0.00 (y n)

ortho位に置換基

これだけでは前提条件がややこしく理解不能!
 精練システムを使って構造を見ると

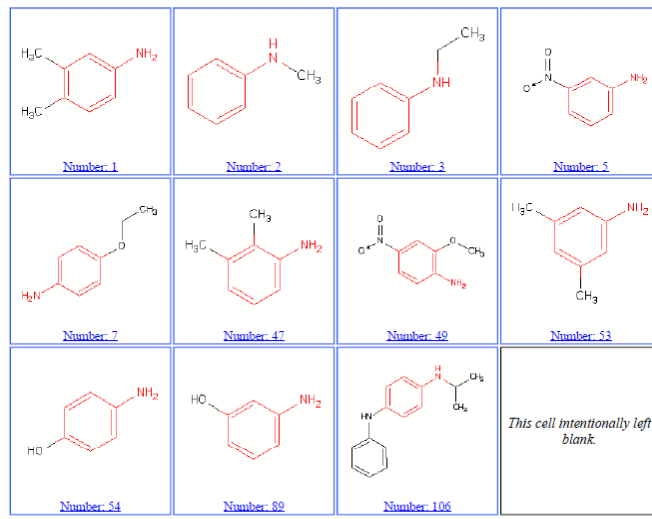
相対重量も増加



この様にして化学者がBAS候補を精査していく。
 この例はGOT減少でもありBASとして不採用。

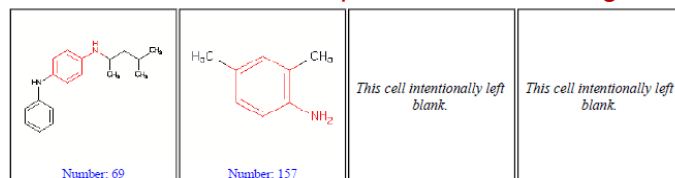
Supporting Structures Chart of Aniline

ヘモグロビン数減少化合物 compounds showing hemoglobin decrease



芳香族アミン
をBASに採用

ヘモグロビン数不変化合物 compounds with no changes in hemoglobin



ラットの28日間反復投与毒性

年度内に以下の毒性を対象とした知識ベース作成

- | | |
|---|--|
| <ul style="list-style-type: none"> ■ 肝毒性
liver toxicity <ul style="list-style-type: none"> □ 肝重量 (LIVER_MR/FR) □ GOT_MR/FR □ GPT_MR/FR | 解析中
still on going |
| <ul style="list-style-type: none"> ■ 腎毒性
kidney toxicity | 解析中
working |
| <ul style="list-style-type: none"> ■ 溶血性貧血毒性 <ul style="list-style-type: none"> □ 赤血球 (RBC_MR/FR)
red blood cell count depression □ ヘモグロビン (HGB_MR/FR)
hemoglobin □ ヘマトクリット (HCT_MR/FR)
hematocrit □ 網状赤血球 (RC_MR/FR)
reticulocytosis | 一応完了 BASiCで公開

ただし、Endpoint評価でプロジェクト
公式見解と差があり、対応を考慮中 |

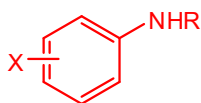
知識ベースの公開

- BASiC知識ベース <http://www.dm-lab.ws/BASiC/>
 - Basic Active Structures in Chemicals
 - 元来, 開発医薬品を対象に開発, 反復投与毒性を追加
 - 現時点では溶血性貧血のみ収載, 肝毒性は資料原簿参照
- 内容
 - ソースデータ & ルール
 - BAS table
 - 構造式, SMARTS検索式, 除外構造, 化合物数
 - 一群のBASとMisc.ですべての対象化合物を覆うことが基本. 支持化合物の重なりはできる限り回避
 - コメント
 - BAS毎の構造式一覧

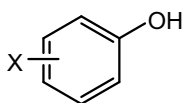
id	Active Structure	Main Condition	SMARTS exception	Inc	Exc	Del	Stp	PDF file
[HGB_FR-dec]		etoccccc1-[NHR]c1	c-[O]([O])c1-c-c1	0	2	1	0	HGB_FR-dec.pdf
[HGB_FR-dB]		etoccccc1-[OH]c1	[-[NH2]]-c-[O]([O])c1-c-[Br].[O]c1c2cc3c([N])ccc3cc12	0	12	4	0	HGB_FR-dB.pdf
[HGB_FR-dC]		[c1cc]c-[O]([O])c1-c2c2-[OH]	None	0	0	0	0	HGB_FR-dC.pdf
[HGB_FR-dD]		C=O	None	1	2	0	0	HGB_FR-dD.pdf
[HGB_FR-dE]		***	None	1	0	0	0	HGB_FR-dE.pdf

雌ラットでヘモグロビンを減少させる4種のBAS

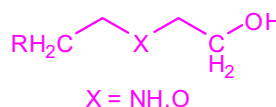
HGB_FR-dec (29)



HGB_FR-dA (0,2,11)



HGB_FR-dB (0,12,4)



HGB_FR-dC (0,0,3)
X = NH, O

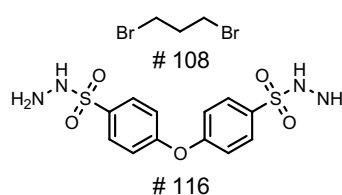
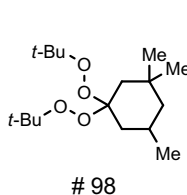
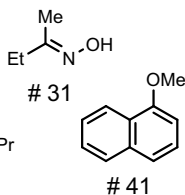
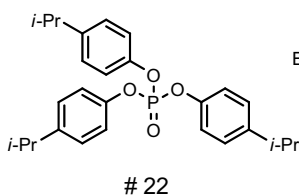


HGB_FRdD (1,26,5)

(l, m, n) は(増加, 無変化, 減少) 化合物数.

グリコールエーテルの毒性は毒性知識情報グループで独立に確認された

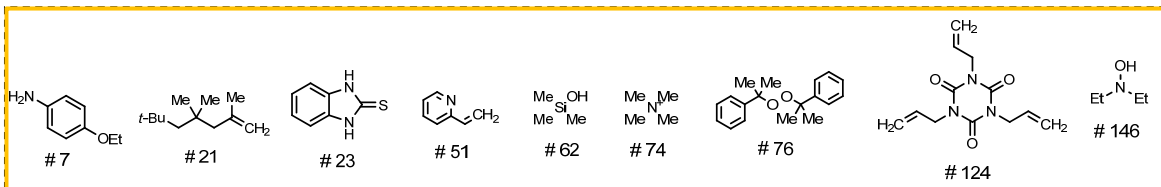
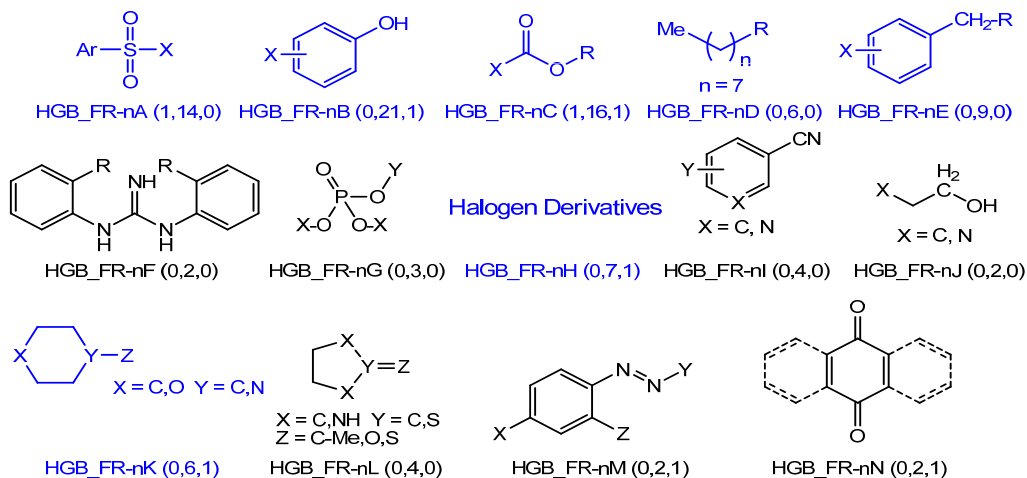
dBとdDのBASは溶血性貧血に関連する場合もあり?



上記のBASでカバーされない毒性化合物6種.

雌ラットでヘモグロビンを減少させない14種のBAS

HGB_FR-n (109)



上記のBASでカバーされない非毒性化合物9種.

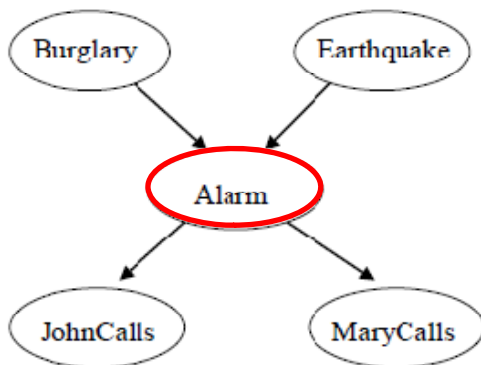
BASiC知識ベース

- 知識の整理にヒントを与える
 - 新しいカテゴリーの模索
 - ベイジアンネットの部分構造ノードとその確率
 - 専門家へのヒント提供
 - 支持構造の提示
- すべて世界でただ1つ
 - 知識ベース自体
 - カスケードモデルと精錬システムによるマイニング法

ベイジアンネットワークとは

- 因果の連鎖を表現
 - 泥棒でアラームは鳴る
 - 地震でアラームは鳴る
 - アラームはマリーに電話をさせる
 - アラームはジョンに電話をさせる

因果の程度を表す条件付確率表 (CPT: Conditional probability table) を作成



B, EからAへの条件付確率表

P(A B,E)	A=T	A=F
B=T, E=T	0.95	0.05
B=T, E=F	0.94	0.06
B=F, E=T	0.29	0.71
B=F, E=F	0.001	0.999

条件

泥棒が侵入して、地震が起きたとき

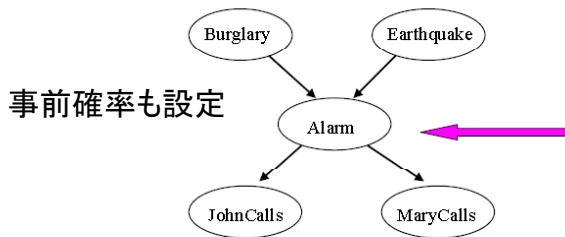
結果

0.95の確率でアラームが鳴る

CPT (conditional probability table)

$P(B) = .001$

$P(E) = .002$ 事前確率



事前確率も設定

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01

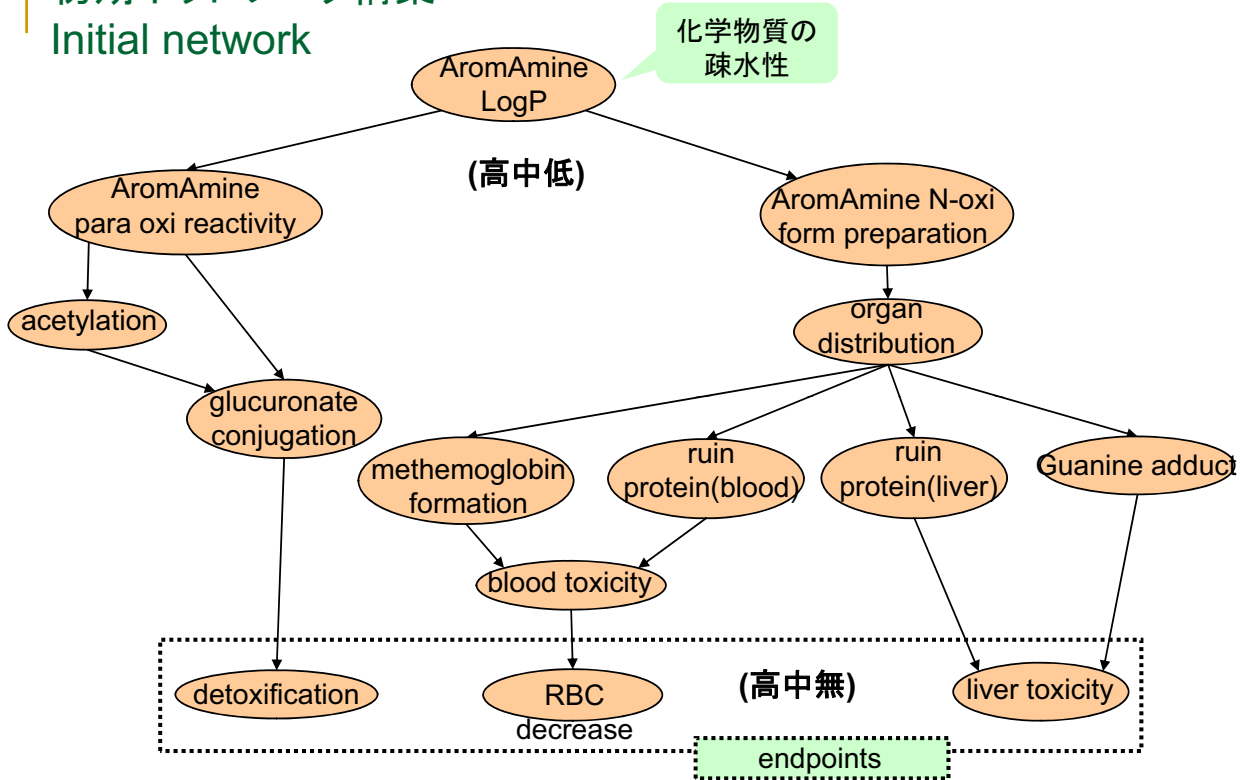
条件付き確率表

エキスパートシステムの現代的健全版

確率推論で何ができる？

- ◆ 「Maryが電話してきた」というevidenceがある時、泥棒の確率は？
 $P(\text{Burglary}|\text{Mary_calls}) = 0.056$ 事前確率の影響で意外と低い
- ◆ 観測データからのEMアルゴリズムによるCPT値の学習

初期ネットワーク構築 Initial network



矛盾解析による専門家からの暗黙知獲得

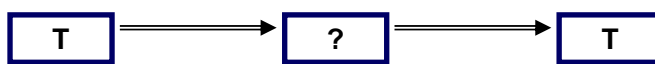
仮定するネット

	T	F
T	.90	.10
F	.10	.90

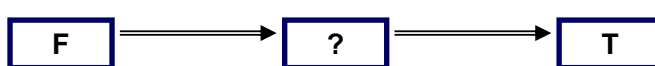
	T	F
T	.80	.20
F	.20	.80



A, Cにevidenceが与えられた時,



納得, 納得
Plausible path



ちょっとおかしい
Conflicting path

Conflict measureを計算し矛盾していれば 1. evidenceが信用できない
2. ネットがおかしい

毒性専門家の頭脳を刺激して, 仮説を引き出す

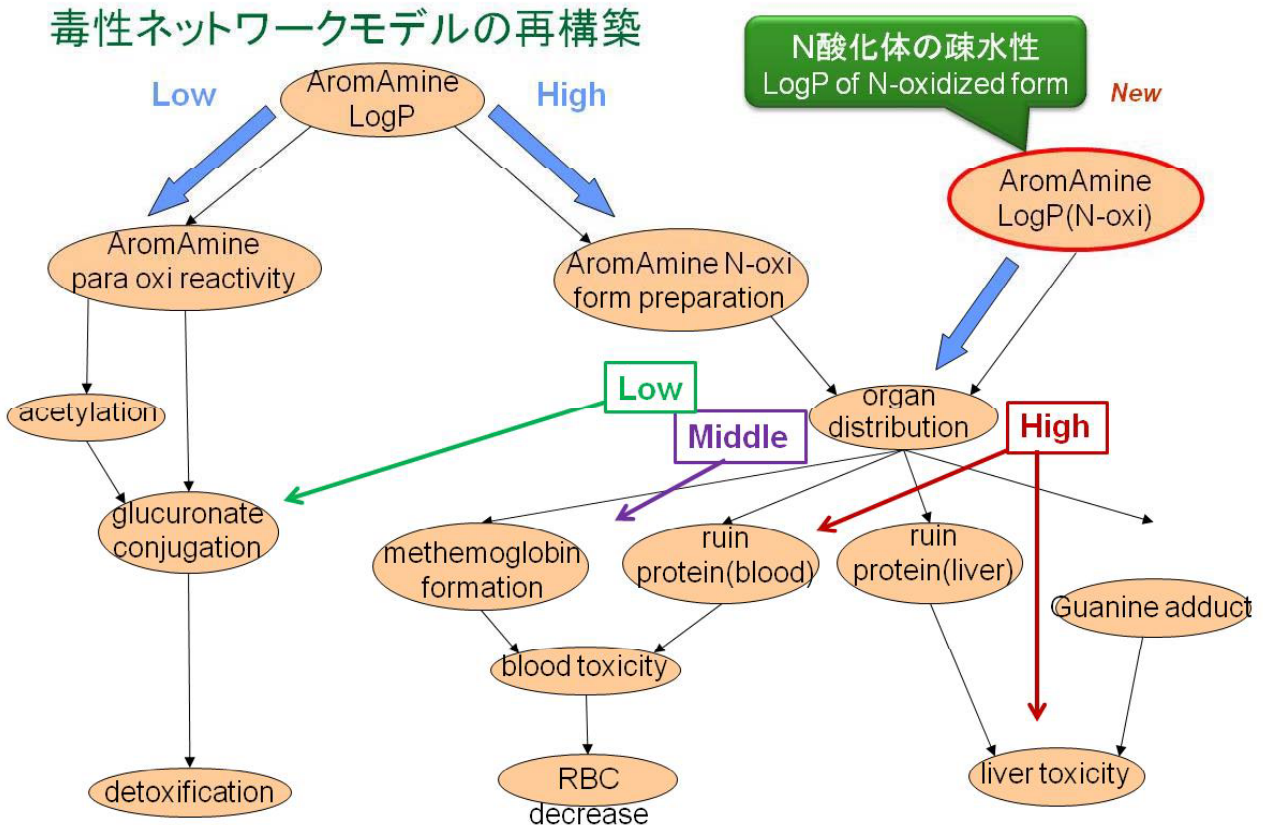
毒性推論実験

- 矛盾を起こした3物質 (芳香族アミン14物資中)
どのようにnetを変更しても、いずれかの化合物で矛盾が起こる。

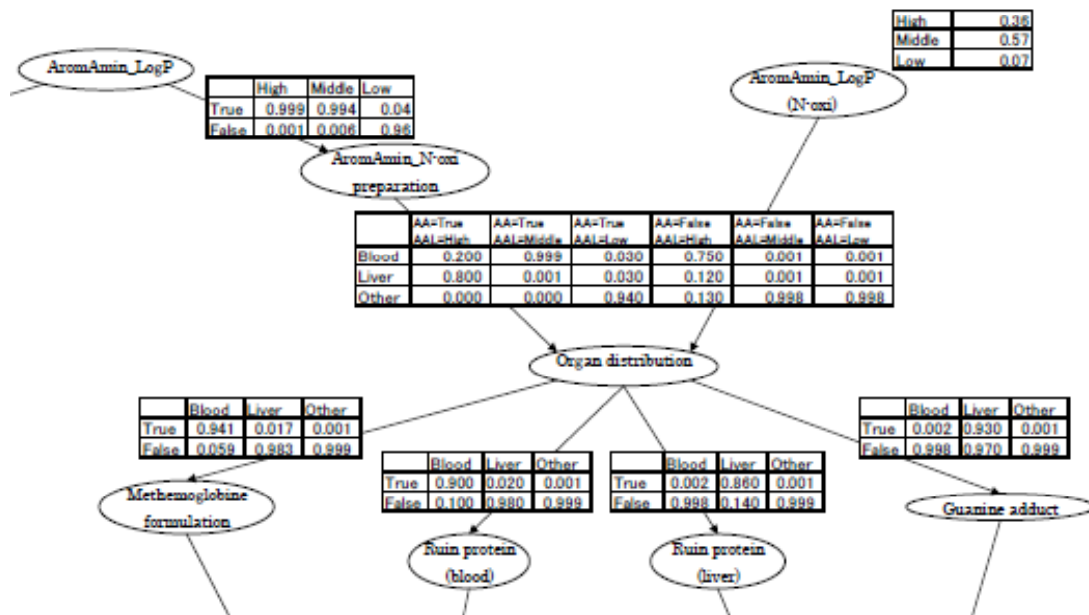
Compounds	Evidences		
	AromAmine LogP	RBC decrease	Liver toxicity
3,4-Dimethylaniline	High	-2.4 ~ 0	High
2,3-Dimethylaniline	High	-4.8 ~ -2.4	Low
3-Nitrobenzenamine	Low	~ -16.8	High

➡ 新たな要因の追加が必要である

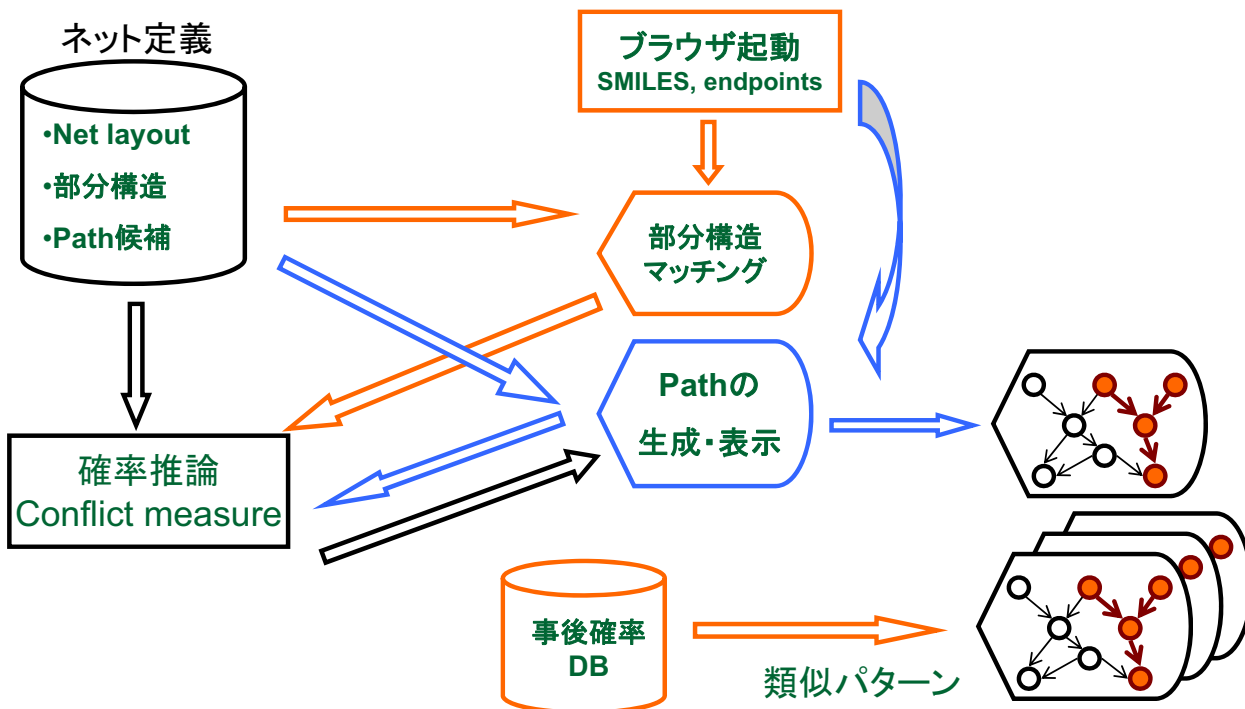
毒性ネットワークモデルの再構築



EMアルゴリズムによる条件付き確率表の推定結果(一部)



ベイジアンネットによる毒性評価システム



現状での画面表示： 予測結果

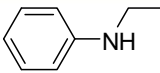
Input SMILES
Input Fragment

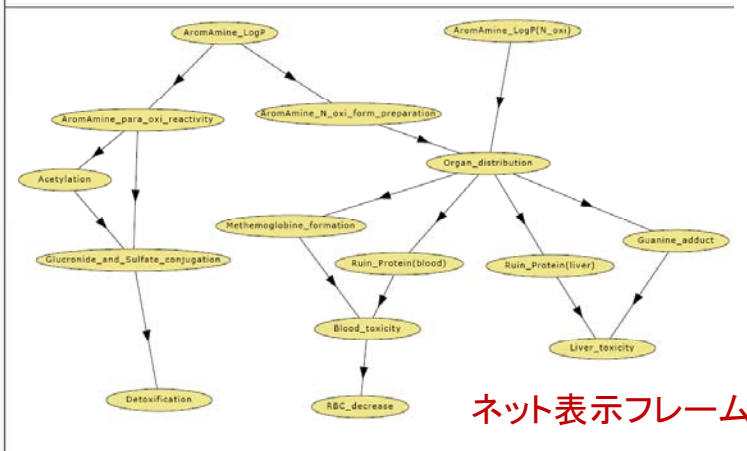
Submit Query

Evidences

<input type="checkbox"/> AromAmine_LogP	High
<input type="checkbox"/> AromAmine_LogP(N_oxi)	Middle
<input type="checkbox"/> Detoxification	False
<input type="checkbox"/> RBC_decrease	High
<input type="checkbox"/> Liver_toxicity	False

ConflictAnalysis

SMILES: CCNc1ccccc1  化合物表示フレーム



ネット表示フレーム

```

----> Thu Jul 9 18:11:11 2009
::: read XML project file :::
read 15 nodes statements

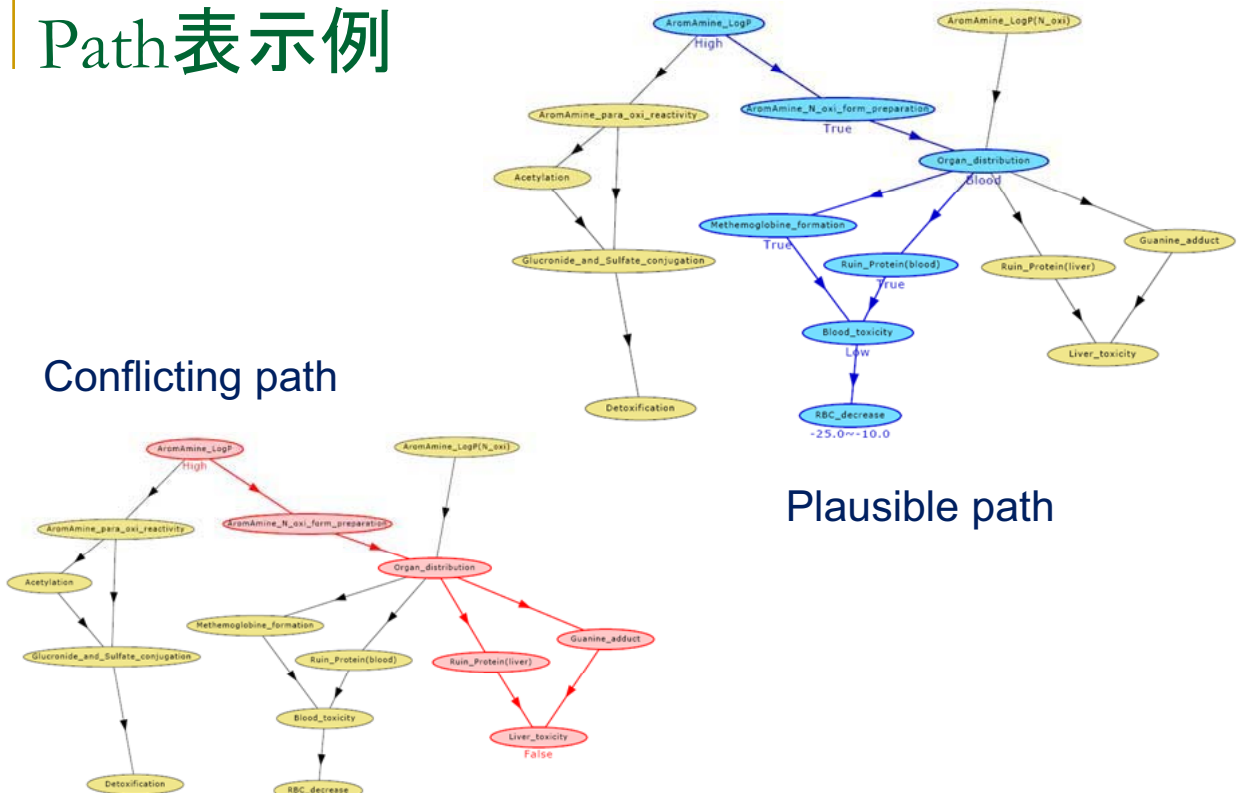
::: set Network Property :::
AromAmine_LogP node size is 3
Blood_toxicity node size is 3
Methemoglobin_formation node size is 2
RBC_decrease node size is 3
    
```

ログ出力フレーム

予測結果

Evidence
入力
予測結果
表示フレーム

Path表示例



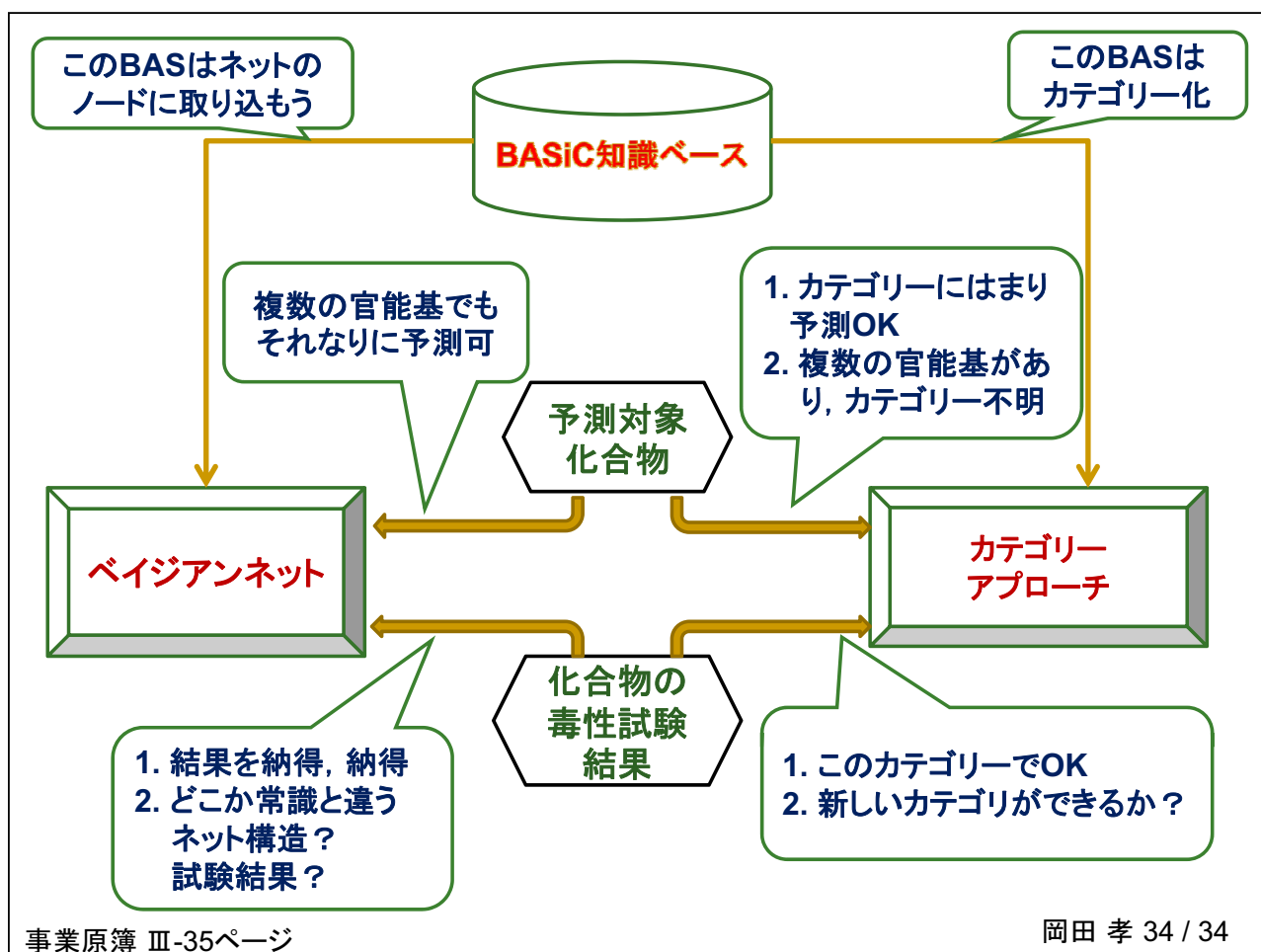
今後の開発予定

- 年度内の開発予定項目
 - 要素技術の統合によるプロトタイプシステムの完成
 - 毒性知識情報データベースから作用機序の組込
 - 拡張したネットと141化合物で条件付き確率の学習
 - 同じく141化合物を対象とした矛盾解析の実行
 - 類似した事後確率分布を示す化合物の提示
- 次年度からの開発予定項目
 - 検知した矛盾の専門家への提示と知識獲得
 - BASiC知識ベース所載部分構造のネットへの取り込み
 - 代謝知識情報データベース内容の取り込み
- システム開発の詳細は事業原簿参照
- 毒性をベイジアンネットで扱うのは世界最初
- 矛盾解析を利用した知識獲得は全分野で日本初

4. 最終目標への課題と 達成見込み

4.最終目標への課題と達成見込み

小項目	最終目標(平成23年度末) 出典:基本計画 p.5	最終目標達成への課題
(2)ベイジアンネットワークを活用した毒性予測機能の開発	毒性知識情報データベース及び代謝知識情報データベースを連携し、毒性予測モデルを組み込み、化学構造を入力すると、代謝物情報や、親物質・代謝物の有害性情報、類似化合物の検索及び解析結果や、肝臓への毒性を中心とした予測結果が表示される有害性評価支援システム統合プラットフォームを開発し、公開する。また、ユーザー独自の情報をデータベースに追加することにより、より精度の向上につながる機能を備えたものとする。開発にあたっては、国際整合性を考慮し、OECD (Q)SAR Application Toolboxとの互換性を確保した仕様とし、カテゴリー化の機能に関しては、OECDで検討されているカテゴリーアプローチの方法論に準拠したものとする。	ベイジアンネットワークによる評価システムは、要素技術の統合自体には問題はないが、実用的なシステムへ発展させるためには、毒性専門家の知識を効果的にネットに反映させることが課題。このような世界初のシステムが完成して評価できること自体は確実である。



実用化の見通し

■ BASiC知識ベース

- 既にInternetで公開. 非専門家も参照.
- 毒性専門家は研究対象の選択に利用
- 内部的にはカテゴリーの構築とベイジアンネットに使用
- 構築用のツール類も, ソースコードをすべて公開予定

■ ベイジアンネット

- プロジェクト終了までにNITEへ移管, 公開可
- 専門家がカテゴリーが不明時の毒性予測に使用
- 専門家が検知された矛盾をヒントとして洞察を深化
- ソースコードをすべて公開予定, ただしベイジアンネットのエンジンとOECHEMライブラリは商用ソフト