

Kumamoto University

2019 - 11 - 25

Computational Toxicology and Chemical Data Science.

Fundamentals and Applications to Drug Discovery

In Silico Data, Ltd.
Kohtaro Yuta

Basic aim of this presentation (1)

1. Discussion of points to consider
when implementing data science in the chemical field
(化学分野でのデータサイエンス実施上での留意点の討論)
2. Discussion on the fusion of chemistry and computer
(化学とコンピューターの融合に関する討論)
3. Discussion of points to consider
regarding the application of data science and AI
(データサイエンスやAI等の適用に関する留意事項の討論)

Basic aim of this presentation (2)

- Aim to be able to discuss correctly and accurately
正しく、且つ精度高く討論できることを目指す
- Technical discussions are not extensive
技術的な基本を広範囲に討論するものではない

自動車を使って目的地に行き、仕事をする
Use your car to go to your destination and work

not a car mechanic

safety drive and get to
your destination reliably,

Turbo, 4WD, inverter, etc.

**Road traffic law, road map,
safety driving, etc.**

Contents:

挨拶: Greetings:

株式会社 インシリコデータ (In Silico Data, Ltd.)

湯田 浩太郎 (Kohtarō Yuta)

◆導入 計算毒性学と「化学データサイエンス」

Introduction: Computational Toxicology and “Chemical Data Science”

◇第一部 計算機化学 (Computer Chemistry) 関連

Part1. Computer Chemistry

◇第二部 化学多変量解析／パターン認識 (ケモメトリックス (Chemometrics)) 関連

Part2. Chemical multivariate analysis / pattern recognition (Chemometrics)

◇第三部 人工知能 (Artificial Intelligence) 関連

Part3. Artificial Intelligence

◇第四部 インシリコ創薬関連

Part4. Insilico drug design

本日のプログラムおよびキーワード:

Contents and Key words (1)

◆導入 計算毒性学と「化学データサイエンス」

Introduction: Computational Toxicology and “Chemical Data Science”

計算毒性学でのコンピューター導入原理、

二大毒性評価関連技術(化学多変量解析/パターン認識アプローチ、人工知能アプローチ)、

データサイエンスから「化学データサイエンス」へ

◇第一部 計算機化学(Computer Chemistry)関連

Computer Chemistry

化合物保存形式、化合物命名法、化合物検索(完全一致、部分構造、2・3次元構造検索、他)手法、

一元一項対応(Canonicalization)、串刺し検索、化合物の扱い(縮合多環、互変異性、立体/幾何異性)、

化合物表記(ケトエノール、ニトロニトロソ、他)

本日のプログラムおよびキーワード:

Contents and Key words (2)

◇第二部 化学多変量解析／パターン認識(ケモメトリックス(Chemometrics))関連

Chemical multivariate analysis / pattern recognition (Chemometrics)

化学パラメーター、2／3次元パラメーター、種々データ解析手法、過剰適合、偶然相関、線形／非線形性、特徴抽出、最少サンプル数、最少パラメーター数、クラスポピュレーション、次元変換／圧縮／縮小、分類率／予測率、要因解析、オートスケーリング、アウトライヤー／インライヤー、解析信頼性指標(サンプル数／パラメーター数)、KY(K-step Yard sampling)法、パーセプトロン、バックプロパゲーション、遺伝的アルゴリズム、ファジー理論、内挿／外挿問題、他

◇第三部 人工知能(Artificial Intelligence)関連

Artificial Intelligence

人工知能の歴史、ルールベース型人工知能、ニューラルネットワーク型人工知能、深層学習、サンプル数問題、要因説明問題、ルールのコンピューターへの組み込み、ネットワーク構造、LISP、FORTRAN、PYTHON、他

◇第四部 インシリコ創薬関連

Insilico drug design

QSAR, ケモメトリックスによる薬理活性／毒性／物性／他、リード化合物検索、リード化合物再構築、インテグレートッド概念、並列創薬、人工知能によるリストラクチャリング、他

In Silico Data, Ltd. Kohtaro Yuta

Educational Background:

Tohoku University (Sendai, Japan)
Pharmaceutical department, Takano research laboratory
Organic synthesis of natural products
March 1979, Dr. of pharmacy from Tohoku university

Working Experiences:

Pennsylvania State University (Pennsylvania, U.S.A.)
Chemistry department, Jurs research laboratory
Computer chemistry and toxicity screening by computer
July 1979 – August 1981 Research associate

Toyohashi University of Technology (Toyohashi, Japan)
Chemistry department, Sasaki research laboratory
Computer chemistry and $^1\text{H-NMR}$ data base development
August 1981 – December 1983 Research assistant

Fujitsu Ltd. (Tokyo, Japan)
System engineer for chemical related research systems
December 1983 – June 2009

National Institute for Environmental Studies
The Center for Environmental Risk Research
July 2009 – March 2010 Executive researcher

Insilico Data Ltd. (Chiba, Japan)
President
May 2010 – current

◆富士通におけるコンピュータケミストリ(計算機化学)関連の業務内容 Computer Chemistry fields

1. 化学関連システムの開発／設計／販売／教育／メンテナンス実施および支援業務

- ・CADD (Computer Aided Drug Design) 関連システム
- ・ケモメトリクスおよび構造－活性相関支援システム
- ・化合物／反応等のデータベースシステム

2. データ解析支援業務

- ・ADAPTおよびSARDシステムを用いた構造－活性相関解析支援
- ・ADAPT、T-SARシステムを用いたケモメトリクス解析支援

3. コンサルタント関連業務

- ・計算機化学関連一般教育
- ・コンピュータによる構造－活性相関教育／解析実施
- ・ケモメトリクス教育／解析実施

4. 情報収集業務

- ・コンピュータケミストリ関連の全般的情報収集
- ・当社製品の競合製品調査

◆化学関連支援システムの設計・開発および支援実績；

Chemistry Related System Development

1. パターン認識による構造－活性相関・ケモメトリクス支援システム；**ADAPT**
(米国ペンシルバニア州立大学(基本部分構築)、富士通(GUI部分構築))
2. 人工知能によるバイオアナログウス化合物創出システム 設計支援と立ち上げ；
EMIL(京都大学農学部との共同開発)
3. 生体内代謝予測支援システム構築(北里大学薬学部との共同開発)
4. 合成ルート創出支援システム構築；**CASINO**(通産省物質研究所との共同開発)
5. 化学スプレッドシートシステム設計支援；**KAMELO**(田辺製薬様との共同開発)
6. 化学ツールキット構築；**CHECK**(富士通開発)
7. 3次元構造－活性相関支援システム構築；**SARD**(富士通開発)
8. JOISオンラインシステムの化合物データ移行設計支援(JOIS;JST(旧JICST))
9. 化学品検査協会のWEB上での毒性予測システム設計支援

◆バイオ関連支援システムの設計・開発および支援実績；

Biotechnology Related System Development

1. モチーフ解析システム設計(国立遺伝学研究所)；初期の参加のみ
2. SNPs解析支援システム
 - ①徳島大学医学部ゲノム解析センター
センター長 板倉光夫教授との共同研究実施(2000年度～)
 - ②東京女子医科大学膠原病リウマチ痛風センター
センター長 鎌谷直之教授との共同研究(2000年度～)

□ 計算毒性学と化学データサイエンス(1)

1. 計算毒性学でのコンピューター導入基本技術

二大毒性評価関連技術

- ① 化学多変量解析 / パターン認識アプローチ
- ② 人工知能アプローチ

1. Computer introduction basic technology in computational toxicology

Two major toxicity evaluation related technologies

- ① Chemical multivariate analysis / pattern recognition approach
- ② Artificial intelligence (Rule base) approach

□計算毒性学と化学データサイエンス(2)

2. データサイエンスから「化学データサイエンス」へ

データサイエンスはデータ処理の汎用的な呼び方である。

適用分野により様々な関連技術が展開される。

例えば、化学分野での多変量解析／パターン認識は

「ケモメトリックス」と呼ばれている。

化学分野でのデータサイエンスは

「化学データサイエンス」

2. From **data science** to **chemical data science**

Data science is a general term for data processing.

Various related technologies will be developed depending on the application field.

For example, multivariate analysis / pattern recognition in the chemical field

It is called “Chemometrics”.

Data science in the chemical field

“**Chemical Data Science**”

□データサイエンスから「化学データサイエンス」へ

From data science to “chemical data science”

- ・化学分野でのデータサイエンスは
「**化学**データサイエンス」 “Chemical Data Science”
- ・創薬分野でのデータサイエンスは
「**創薬**データサイエンス」 "Drug discovery data science"
- ・毒性分野でのデータサイエンスは
「**毒性**データサイエンス」 “Toxicology Data Science”
- ・物性分野でのデータサイエンスは
「**物性**(マテリアルズ)データサイエンス」“Materials Data Science”
- ・その他

1. 計算毒性学でのコンピューター導入原理 (1)

□化合物毒性の特徴とコンピューター導入の困難性

Characteristics of compound toxicity and difficulty in introducing computer

①メカニズム不明／複雑：メカニズム中心の議論困難

Mechanism unknown / complex: mechanism-centered discussion difficult

発現メカニズム等複雑で、毒性の種類も多い

Complex expression mechanism and many toxic types

1. 計算毒性学でのコンピューター導入原理 (2)

②化合物構造変化性極大: 構造類似性による情報適用困難

Compound structure variability maximum:
difficult to apply information due to structural similarity

メタンレベルからマクロライドまで、全化合物が評価対象

All compounds are evaluated from methane level to macrolide

③高い予測率が要求される: 予測結果重視の傾向

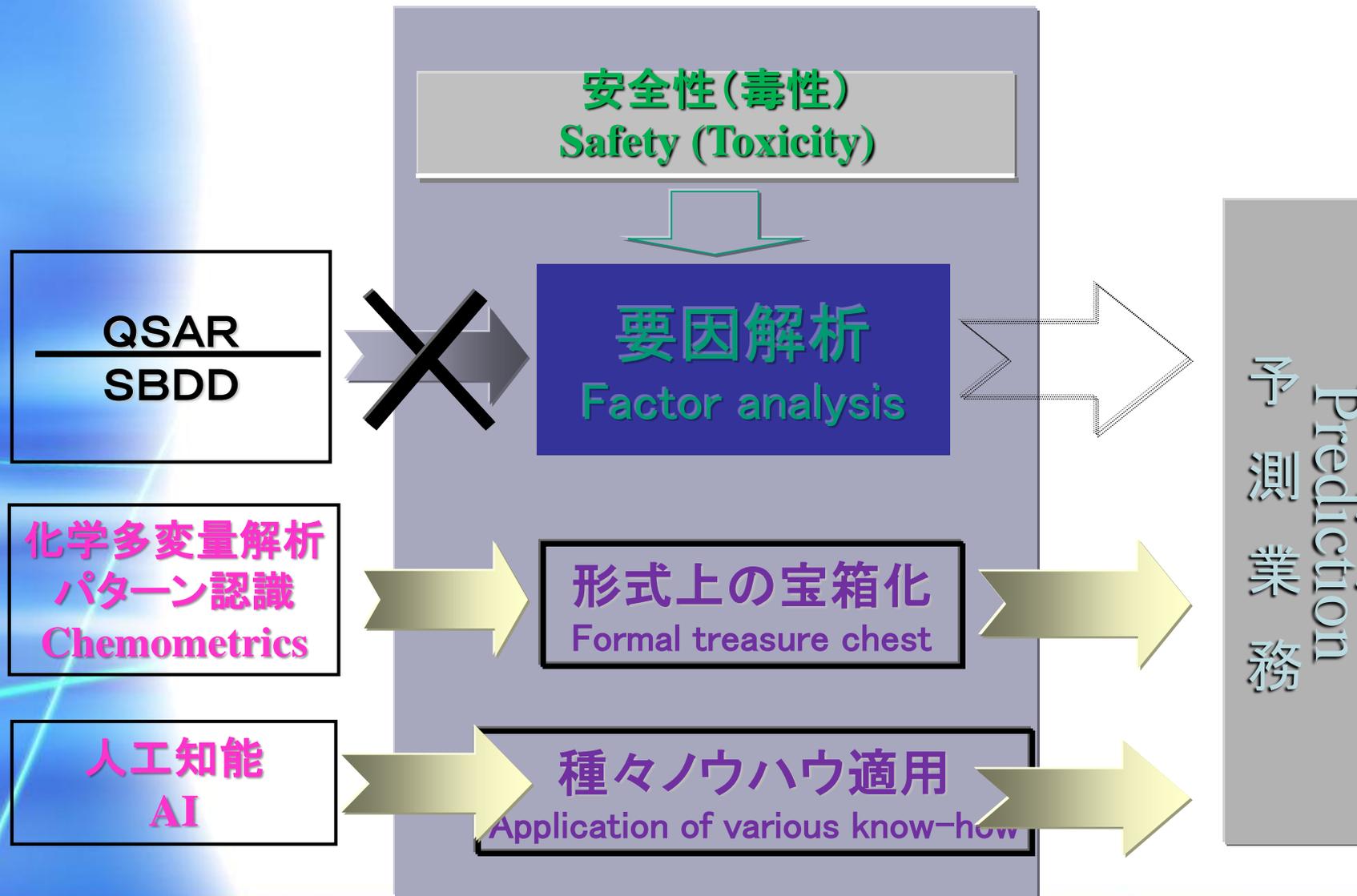
A high prediction rate is required: Trends that place importance on prediction results

毒性予測は失敗が許されないなので、要求レベルが高い

Toxic prediction is not allowed to fail, so demand level is high

コンピュータによる安全性予測へのアプローチ

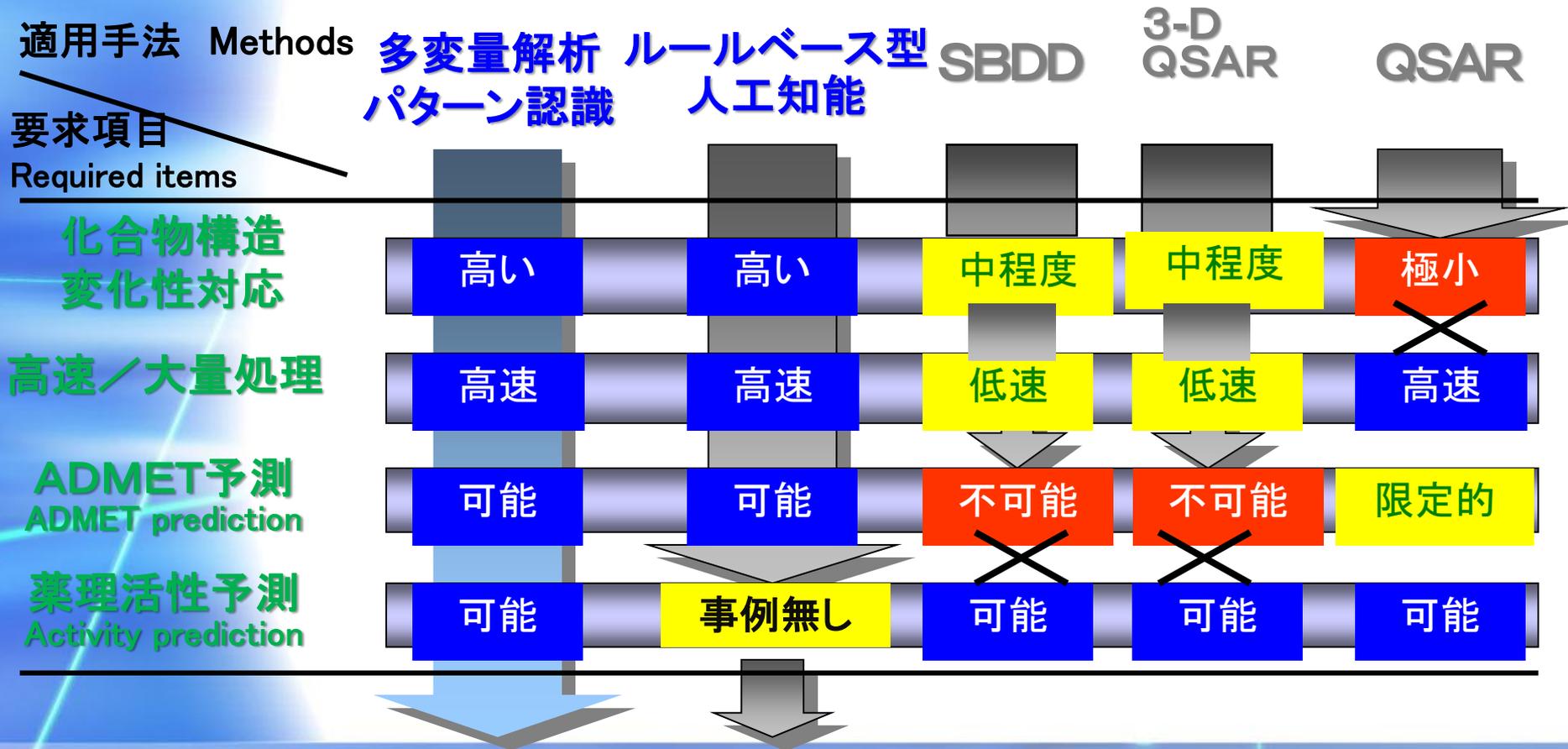
Approach to safety prediction by computer



□要求項目と適用手法の関係(1)

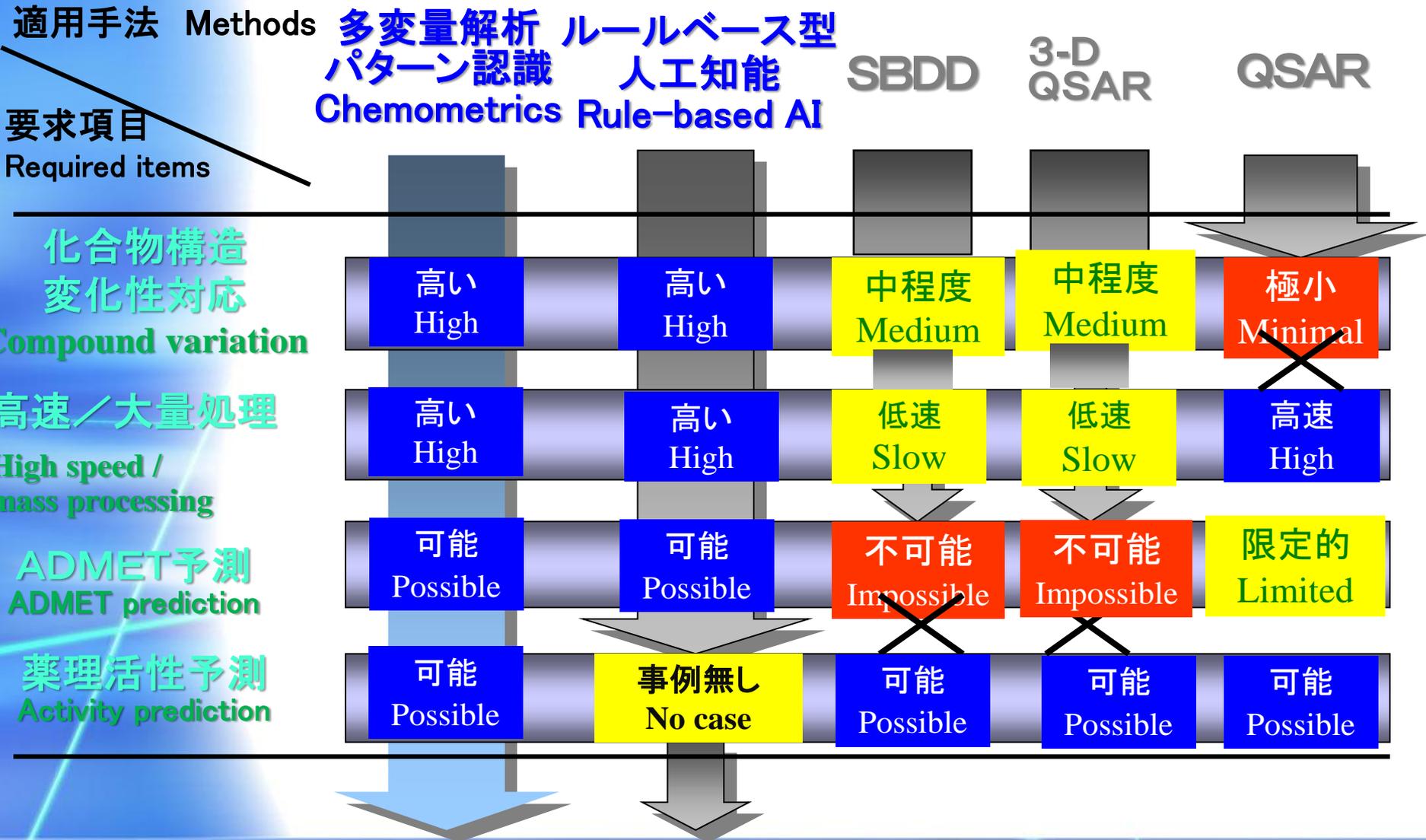
Relationship between requirement items and applicable methods

安全性(毒性)、薬理活性、ADME予測上での解決すべき要求項目と適用可能アプローチ
 Safety (toxicity), pharmacological activity, requirements to be solved in ADME prediction and applicable approaches



□要求項目と適用手法の関係(2)

Relationship between requirement items and applicable methods



□ 二大毒性評価関連技術

Two major toxicity assessment related technologies

毒性の特性より、毒性評価に適用可能なアプローチは適用内容や基本原理より以下の二種類となる

Based on the characteristics of toxicity, there are two approaches that can be applied to toxicity assessment based on the contents of application and basic principles.

① 化学多変量解析／パターン認識アプローチ、

Chemical multivariate analysis / pattern recognition approach

② 人工知能アプローチ Artificial intelligence approach

□ ICH-M7では、医薬品に含まれる不純物の遺伝毒性（変異原性）安全性評価に互いに相補的な

① データ解析アプローチと ② 人工知能アプローチの両方を採用するように求めている。

ICH-M7 requires that both (1) data analysis approach and (2) artificial intelligence approach, which are complementary to each other, be evaluated for genotoxicity (mutagenicity) safety assessment of impurities contained in pharmaceutical products.

* 結果が矛盾する場合は、専門家によるレビューを実施

* If the results are inconsistent, review by experts

□適用分野単位のデータサイエンス

Data science for each application field

<u>基本技術</u>	<u>適用分野</u>	<u>名称</u>
Basic technology	Application fields	Name
<p>データサイエンス Data science</p>	<p>化学 Chemistry 創薬 Drug design 毒性 Toxicity 物性 Physical properties バイオ Bio- 蛋白 Protein 医療 Medical care 環境 Environment その他 Others</p>	<p>データサイエンス Data science</p>
<p>多変量解析 パターン認識 Multivariate analysis / pattern recognition</p>		
<p>AI Artificial intelligence</p>		