Chapter 6

In Silico Methods for Carcinogenicity Assessment

Azadi Golbamaki and Emilio Benfenati

Abstract

Screening compounds for potential carcinogenicity is of major importance for prevention of environmentally induced cancers. A large sequence of alternative predictive models, ranging from short-term biological assays (e.g. mutagenicity tests) to theoretical models, have been attempted in this field. Theoretical approaches such as (Q)SAR are highly desirable for identifying carcinogens, since they actively promote the replacement, reduction, and refinement of animal tests. This chapter reports and describes some of the most noted (Q)SAR models based on the human expert knowledge and statistically approach, aiming at predicting the carcinogenicity of chemicals. Additionally, the performance of the selected models has been evaluated and the results are interpreted in details by applying these prediction models to some pharmaceutical molecules.

Key words Carcinogenicity, Structural alerts, Genotoxicity, Non-genotoxicity, QSAR, In silico, Toxtree, SARpy, Applicability domain index

1 Introduction

The study of the chemical carcinogenesis mechanisms and determining the safety of the existing and the new chemicals are of increasing importance and necessity to protect human health. From the point of view of mechanism of action, the carcinogens are classified into: (a) genotoxic carcinogens, which cause damage to DNA—many known mutagens are in this category, and often mutation is one of the first steps in the development of cancer [1]; and (b) epigenetic or non-genotoxic carcinogens that do not bind covalently to DNA, do not directly cause DNA damage, and are usually negative in the standard mutagenicity assays [2]. The unifying feature of all genotoxic carcinogens is that they are either electrophiles or can be activated to electrophilic reactive intermediates. This fact has been originally proposed by the Miller's [3, 4]. On the contrary, non-genotoxic carcinogens act through a large variety of different and specific mechanisms.

The mechanisms of action and the metabolic fate of a large number of carcinogens have been already investigated. These studies shed light on the structural features that were frequently present in carcinogenic compounds. Several chemical functional groups and structural alerts (SAs) were identified by researchers through analysis of the results of experimental (veterinary laboratory) carcinogenicity tests. These compounds were mainly genotoxic carcinogens as supported by the specific results from tests for genotoxicity (Ames test [5], Micronucleus assay [6], etc.). Diversely, the recognition of SAs for non-genotoxic carcinogens is far behind, because no unifying theory provides scientific support. A number of SAs and characteristics of several types of non-genotoxic carcinogens have been summarized by Woo et al. [2] (*see* **Notes 1** and **2**).

The long-term carcinogenesis bioassays using animal testing methods have played a central role in assessment of chemical's carcinogenicity, however, for ethical and practical reasons their use is dramatically declining, and the genotoxicity short-term tests have taken the pivotal role in the pre-screening of carcinogenicity. The need to reduce animal testing, time, and cost in the process of assessment of carcinogenicity of chemicals had lead to an increased use of in silico methods as toxicological risk assessment tools. Among the in silico methods, the use of (Q)SAR models is supported by several legislative authorities (REACH [7]) upon fulfillment of the required characteristics of a (Q)SAR model according to the indications reported by different legislations. This goes hand in hand with the progress made in the field of the computational predictive models to date.

(Q)SARs are often incorporated into expert systems. An expert system is any formalized system that is mostly computer-based, and that can be used to make predictions based on prior information [8].

There are many (Q)SAR models published in the literature for predicting genotoxicity and carcinogenicity. The most commonly modeled endpoint for genotoxicity is the Ames test mutagenicity. The application of the Ames test to large numbers of chemicals has shown that this test has a high predictivity for chemical carcinogens (around 80 %) [9]. Most models are classifiers that predict a chemical compound as genotoxic (and thus carcinogenic) or not. Since the recognition of non-genotoxic carcinogenicity SAs is not extended compared to genotoxic SAs, few models are available for identifying non-genotoxic carcinogens [10]. While the SAs for genotoxic carcinogens have been identified to a high extent and used widely within predictive models for genotoxicity, the SAs for identifying non-genotoxic carcinogens are still a concern for the investigators. Benigni et al. (Toxtree 2.6.0) have recently enhanced the set of non-genotoxic SAs that captures carcinogens [9]. This list can provide a considerable insight to the possible variety of

mechanism of actions underlying the non-genotoxic carcinogenicity. Hence, the approaches for (Q)SAR analysis and identification of SAs for non-genotoxic carcinogens differ accordingly to their specific mechanism of action of these chemicals (interaction with proteins, DNA replication enzymes, etc.) (*see* **Note 1**). A number of SAs and characteristics of several types of non-genotoxic carcinogens have been summarized and discussed by Woo et al. [2].

However, statistical-based models will provide predictions that are based on the knowledge acquired from the training set that had been used to develop the model. In fact, these models are suitable in predicting both genotoxic and non-genotoxic carcinogens. For unknown non-genotoxic SAs, the statistical-based models can fill the information gap. In other words, these models may provide insight into the recognition of the missing information in the SAs list developed by human experts by investigation through experimental results mostly based on the Ames test.

In the context of prediction of carcinogenicity by (Q)SAR models, it is essential to integrate results from both expert systems and statistical-based models. This approach will considerably improve the prediction performance of (Q)SARs.

There are several commercial and non-commercial expert systems for predicting genotoxicity and carcinogenicity [11, 12]. Freely available models include VEGA-CAESAR [13], SARpy [14], Toxtree [15], OncoLogic [16], OECDE Toolbox [17], and lazar [18]. Alternately, MultiCASE [19], TOPKAT [20], HazardExpert [21], and DEREK [22, 23] are some of the most common commercial expert system.

Expert systems are based on three main modeling approaches which are rule-based, statistical-based, or hybrid methods [24]. Rule-based methods codify the human rules which identify certain potential molecular fragments responsible for carcinogenicity. Statistical models extract the information from a set of chemicals by using data mining methods [25].

Rule-based systems combine toxicological knowledge, expert judgment, and fuzzy logic. OncoLogic, DEREK, HazardExpert as well as implemented modules in Toxtree and the OECD Toolbox are rule-based systems.

Statistical-based systems use a variety of statistical, rule-induction, artificial intelligence, and pattern recognition techniques to build models from different databases used as training sets. For example, MultiCASE and TOPKAT are commercial statistical-based models while lazar and VEGA-CAESAR are statistical-based and publicly available. Additionally, most of the models published in the literature but not implemented are statistical-based (*see* **Note 2**).

A description of some of the most common non-commercial (Q)SAR models is provided below. Three case studies are given in this chapter to illustrate the use and the performance of a number of these models.

2   QSAR Models for Carcinogenicity

2.1   VEGA-CAESAR
(Version 1.1.0)

CAESAR is a model implemented in the VEGA platform [26]. This model uses a statistical-based approach to generate categorical carcinogenicity models. CAESAR is based on the counter-propagation artificial neural network (CP ANN) algorithm. Artificial neural networks (ANNs) as a statistical approach appear to be suitable and promising for prediction of carcinogenicity for dissimilar data sets of chemicals. One of the main advantages of ANNs is that non-linear relationships can be modeled without any assumptions about the form of the model.

2.2   Toxtree
(Version 2.6.0)

Toxtree is a standalone expert rule-based SAR program. This application is a classifier that places chemicals into categories and predicts various kinds of toxic effect by applying decision tree approaches, including the Begnini-Bossa rule-base for mutagenicity and carcinogenicity [27]. The Toxtree module applies human expert rules developed by Begnini and Bossa to identify SAs for mutagenicity and carcinogenicity that may be present in a chemical structure. Carcinogenic SAs are functional groups or substructures that are mechanistically and/or statistically associated with the induction of cancer. Begnini-Bossa SAs for the prediction of mutagenicity and carcinogenicity are highly correlated with Ames mutagenicity. The Begnini-Bossa system contains a list of SAs for the evaluation of carcinogenicity. Structural features represented in the system are easy to understand and interpretable since they have a mechanistic foundation. Toxtree offers additional QSAR models for aromatic amines and alpha, beta-unsaturated aldehydes. The Toxtree output contains "structural alert for genotoxic carcinogenicity" that shows the presence or absence of a SA for Salmonella mutagenicity, and "structural alert for non-genotoxic carcinogenicity" that indicates the presence or absence of a non-genotoxic (epigenetic) SA.

2.3   SARpy
(Version 1.0)

SARpy is a desktop software based on a statistical modeling approach. Through a data mining method, SARpy extracts relevant fragments (molecular substructures) from the analysis of the correlation between the structure, written with simplified molecular input line entry system (SMILES) format, and the endpoint. Using SARpy, and a data set of chemicals with valid experimental results (binary categorical data), users can develop new classification models. SARpy is able to extract both "ACTIVE" (e.g. carcinogenic) and "INACTIVE" (e.g. non-carcinogenic) fragments from chemical structures. In order to discover new carcinogenic SA, we combined three different carcinogenesis databases as a training set and by the aid of SARpy, developed a new carcinogenicity model which consists of a rule set or a collection of SMARTS with their likelihood ratio values in the mentioned training set.

The data gathered for the development of this new rule set are carcinogenicity data collections based on studies on different species. In particular, the data in the training set are a combination of: (1) the carcinogenicity data set (rat) of the EU-funded ANTARES project [28]; (2) the long-term carcinogenicity bioassay on rodents (rat and mouse) ISSCAN data set [29]; and (3) the carcinogenicity (rat and mouse) data set provided by Kirkland et al. [30]. The data set (1680 chemicals together with their carcinogenicity data) built as described above was used as the training set for the extraction of rules. SARpy extracted more than 100 rules from which by applying a human expert judgment we selected 130 rules. The human expert selection aimed to delete the alerts that produced a high number of false negative or false positive predictions. The performance of this model, as tested on the test set obtained from eChemPortal inventory (258 compounds), was as follows: accuracy = 0.67, sensitivity = 0.62, specificity = 0.70.

2.4 OncoLogic™ (Version 8.0)

OncoLogic™ [31] is a desktop computer program released by the U.S. Environmental Protection Agency (EPA) [32] that evaluates the likelihood that a chemical may cause cancer. OncoLogic™ predicts cancer-causing potential by: applying the rules of structure–activity relationship (SAR) analysis, mimicking the decision logic of human experts, and incorporating knowledge of how chemicals cause cancer in animals and humans. This version of the software has a new CAS/name look-up feature under the "Organics SAR" module for approximately 1500 chemicals for which available cancer data can be used directly to create a chemical report. This removes the need to draw the chemical structure for these substances as was necessary in the previous versions of the software.

2.5 Lazar

Lazy structure–activity relationships (lazar) [18] is a standalone program with k-nearest-neighbor approach which can predict chemical endpoints from a training set based on structural fragments. It uses a SMILES file and precomputed fragments with occurrences as well as target class information for each compound as training input. It also features regression, in which case the target activities consist of continuous values. Lazar uses activity-specific similarity (i.e. each fragment contributes with its significance for the target activity) that is the basis for predictions and confidence index for every single prediction.

3 Case Studies

3.1 Case Study 1

An example of Toxtree (v.2.6.0) carcinogenicity prediction.

As it is explained in the Toxtree user manual [33] for estimating carcinogenicity with Toxtree, the following steps should be taken: After launching Toxtree in Windows™ platform, first,

the chemical structures for analysis may be submitted by inserting directly the SMILES, or by using an interactive 2D graphical editor, or in a batch mode by using CSV, TXT, or SDF file formats. Second, among the list of decision tree modules the user may select "carcinogenicity (genotox and non-genotox) and mutagenicity rule-base by ISS" [27] option from the Method menu. Finally, in order to apply the active decision tree on the current compound, the Estimate button should be pressed. If one or more genotoxic or non-genotoxic SA are found in the molecular structure, the name and the identification number of that SA are indicated in the graphical user interface, and the chemical is predicted as carcinogen. Otherwise, the prediction result will be non-carcinogen. Figure 1 shows an example of classification result visualization.

Captafol is an antibacterial drug and fungicide and is categorized as a carcinogen in the Carcinogenic Potency Database (CPDB) [34]. Toxtree v. 2.6.0 finds a SA for genotoxic carcinogenicity (QSA8_gen.Aliphatic halogens) and a SA for non-genotoxic carcinogenicity (QSA50_nogen.dicarboximid) in this chemical structure. By clicking on the name of these two SAs, they become highlighted and the user can see their position in the chemical structure (Fig. 2). The classification results can be saved as a file (CSV, SDF, or TXT format), together with the list of applied SAs.

3.2   Case Study 2   2-Amino-5-nitrothiazole or aminonitrothiazole is an antiprotozoal drug. Antiprotozoal agent is a class of pharmaceuticals used in the treatment of protozoan infection. Figure 3 shows the chemical structure and Table 1 shows the carcinogenicity test summary report as published by the CPDB [34]. Based on the experimental results of TD_{50} on rat species, this chemical is considered as a carcinogen.

VEGA-CAESAR (v. 1.1.0), lazar, Toxtree (v. 2.6.0), and the SARpy (v. 1.0) model predicted this chemical correctly as carcinogen. Figure 4 shows two genotoxic SAs found in the chemical structure of 2-amino-5-nitrothiazole: "SA_27: Nitro-aromatic" and "SA_28: primary aromatic amine, hydroxyl amine and its derived esters". VEGA-CAESAR returned applicability domain (AD) index of 0.5 for the prediction of this drug, and the explanation is "the predicted compound is outside the AD of the model." The "measured activity" of lazar given in the output is "Experimental result(s) from the training data set," so the chemical is inside the AD of the program. Toxtree and SARpy do not report any AD index in their predictions.

Performing prediction with the model constructed by means of SARpy for this chemical, an additional fragment is recognized as responsible for the carcinogenicity property. Figure 5 shows the SA found by this model. Overall, based on these multiple predictions, we can see that there is agreement, even though each model has a different level of reliability.

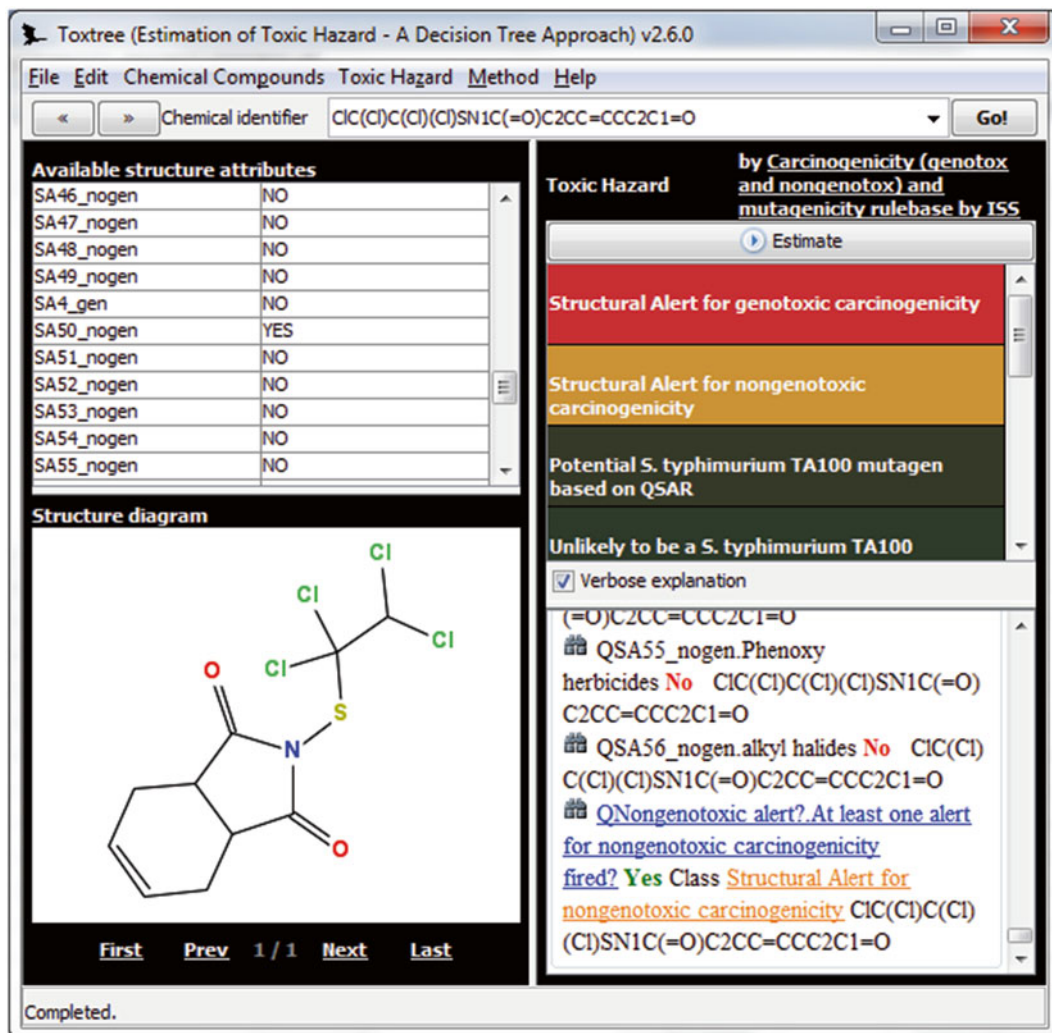As a conclusion, all evidences point toward a carcinogenic effect.

Fig. 1 Toxtree v. 2.6.0 mutagenicity and carcinogenicity prediction for Captafol

3.3 Case Study 3

Bemitradine is an antihypertensive, vasodilator agent, and a diuretic. Figure 6 shows the chemical structure and Table 2 shows the carcinogenicity test summary report as published by the CPDB. Based on the experimental results of TD_{50} on rat species, this chemical is considered as carcinogen.

Toxtree (v. 2.6.0) and SARpy (v. 1.0) model predicted this chemical correctly as carcinogen; conversely, VEGA-CAESAR (v. 1.1.0) and lazar prediction for this chemical was non-carcinogen. Figure 7 shows the genotoxic SA found in the chemical structure, whereas the model constructed by means of SARpy matched another fragment to the molecular structure as responsible for the carcinogenicity property. Figure 8 shows the SA found by the SARpy model. Toxtree and SARpy do not have any
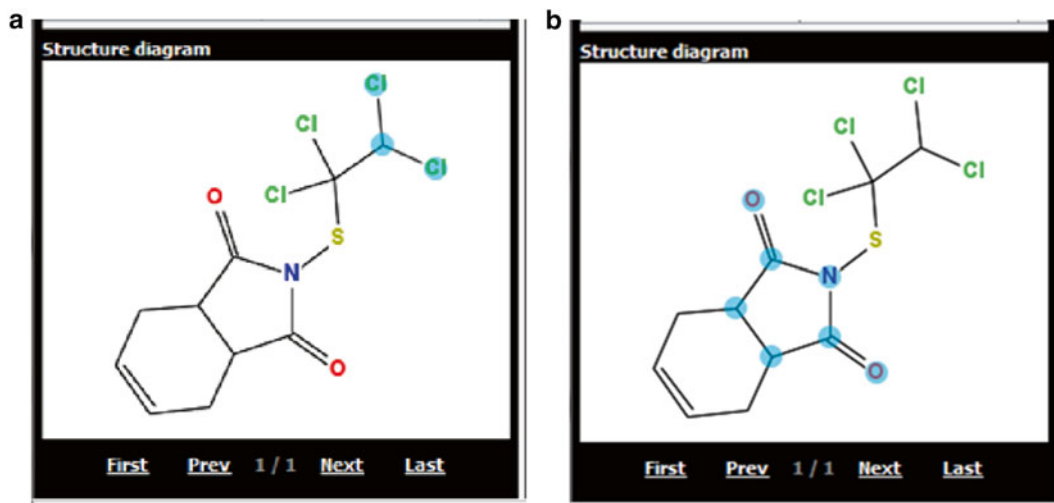
Fig. 2 Genotoxic and non-genotoxic structure alerts found by Toxtree 2.6.0 for Captafol; (**a**) QSA8_gen.Aliphatic halogens; (**b**) QSA50_nogen.dicarboximid are highlighted in the molecular structure
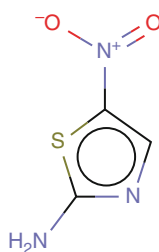

Fig. 3 2-Amino-5-nitrothiazole, with CAS number: 121-66-4 and SMILES: O=[N+]([O−])c1cnc(N)s1

Table 1
Cancer test summary reported in the CPDB for 2-amino-5-nitrothiazole

Rat target sites		Mouse target sites		TD50 (mg/kg/day)	
Male	**Female**	**Male**	**Female**	**Rat**	**Mouse**
No positive	kid lun mgl[a]	No positive	No positive	44.6	No positive

[a]*kid* kidney, *lun* lung, *mgl* mammary gland

AD index along with their prediction results. The AD index of VEGA-CAESAR for this chemical is equal to zero and in the prediction output file it is reported that the predicted compound is outside the AD of the model. The lazar confidence index for its prediction is 0.02.
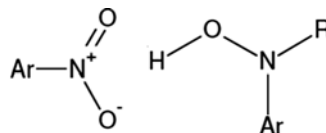
Fig. 4 Genotoxic structure alerts found by Toxtree in the molecular structure of 2-amino-5-nitrothiazole; SA_27: Nitro-aromatic is shown on the *left* side, while SA_28: primary aromatic amine, hydroxyl amine and its derived esters is shown on the *right*, where Ar stands for any aromatic/heteroaromatic ring and R stands for any atom/group
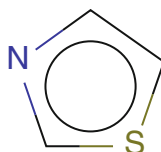

Fig. 5 Carcinogenicity structure alert found by the SARpy model for which the chemical is predicted as carcinogen
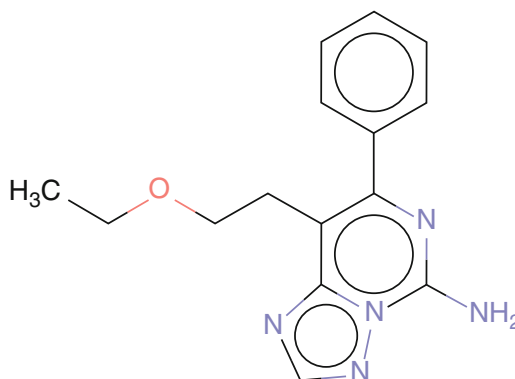

Fig. 6 Bemitradine chemical structure with CAS number: 88133-11-3 and SMILES: n2cnn3c(nc(c1ccccc1)c(c23)CCOCC)N

Table 2
Cancer test summary reported in the CPDB for Bemitradine

Rat target sites		Mouse target sites		TD50 (mg/kg/day)	
Male	**Female**	**Male**	**Female**	**Rat**	**Mouse**
liv	liv mgl[a]	No test	No test	548 m	No test

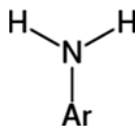[a]liv liver, *mgl* mammary gland

Fig. 7 QSA28_gen. Primary aromatic amine, hydroxyl amine, and its derived structure alert found by Toxtree in the molecular structure of Bemitradine
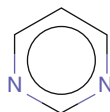

Fig. 8 Carcinogenicity structure alert found by the SARpy model for which the chemical is predicted as carcinogen

Toxtree (v. 2.6.0) prediction for this chemical was: "Negative for non-genotoxic carcinogenicity and positive for genotoxic carcinogenicity." The SA recognized by Toxtree in the molecular structure is "QSA28_gen. Primary aromatic amine, hydroxyl amine and its derived esters (with restrictions)."

However, there are two restrictions to this rule. In fact, if the following conditions are true then the compound is predicted as non-carcinogen:

- Chemicals with *ortho*-disubstitution, or with an ortho carboxylic acid substituent are excluded.
- Chemicals with a sulfonic acid group (–SO₃H) on the same ring of the amino group are excluded.

and in this case study, none of them are applied.

Overall, on the basis of the results of the different models and the low confidence value of lazar and the fact that it is out of AD of VEGA-CAESAR, of course one cannot exclude the possible carcinogenic effect. On the contrary, there are elements to support the toxic effect which cannot be ruled out by the presence of some results going in the opposite direction. Thus, the overall assessment should go for carcinogenicity, but with a higher uncertainty, compared to the results for the case study 1.

3.4   Case Study 4   Amobarbital (formerly known as amylobarbitone or sodium amytal) is a drug that is a barbiturate derivative (*see* Fig. 9). It has sedative-hypnotic properties. On the basis of CPDB it is classified as a non-carcinogen (*see* Table 3). Toxtree (v. 2.6.0), lazar, VEGA-CAESAR (v. 1.1.0), and the SARpy (v. 1.0) model predicted this molecular structure correctly (i.e. non-carcinogen) as confirmed by the experimental result. In addition, the VEGA-CAESAR prediction result has a reliability feature that for this compound: "the predicted compound is into the Applicability Domain of the model."
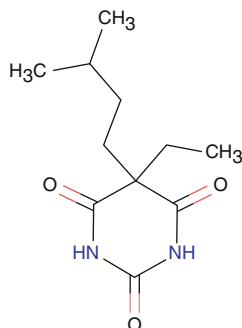
Fig. 9 Amobarbital with CAS number: 57-43-2 and SMILES: CCC1(CCC(C)C)C(=O) NC(=O)NC1=O

Table 3
Cancer test summary reported in the CPDB for Amobarbital

Rat target sites		Mouse target sites		TD50 (mg/kg/day)	
Male	Female	Male	Female	Rat	Mouse
No positive	No test	No test	No test	No positive	No test

In fact, the model has the experimental value of this compound. The AD index of this chemical in the VEGA-CAESAR prediction is equal to 1 (*see* **Note 3**). The lazar reported this chemical as an already existing chemical inside its training set, so we consider it inside its AD. As it is mentioned above, Toxtree and SARpy do not have any AD index along with their prediction results.

As a conclusion, all the prediction results of the above-mentioned models indicate the non-carcinogenic effect of the compound, which are concordant with the experimental value.

4   Notes

1. The different sources of the data used within the different models should always be considered. The CAESAR model is closely related to the rat carcinogenicity, while other models tend to balance results from different studies. There may be differences between the carcinogenicity in animals and in humans [32].

2. It should be noted that the data available for building carcinogenicity models derive studies which identified in several cases effects on different organs (i.e. test for hepatocarcinogenicity, polmonarcarcinogenicity). Therefore, building organ-specific

carcinogenicity may be the best approach in order to obtain models with higher prediction performance. Nevertheless, the number of experimental results on organ-specific carcinogenicity is at the time limited making them inadequate for building a (Q)SAR model with high performance.

3. VEGA provides the experimental result of the target compound, if available. The experimental value prevails on the predicted one, and thus the AD index is 1. The predicted value of the target compound is also given in the summary page.

References

1. Arcos JC (1995) Chemical induction of cancer: modulation and combination effects. An inventory of the many factors which influence carcinogenesis. Springer Science & Business Media, New York

2. Woo Y, Lai D (2003) Mechanisms of action of chemical carcinogens, and their role in Structure-Activity Relationships (SAR) analysis and risk assessment. In: Benigni R (ed) Quantitative Structure-Activity Relationship (QSAR) models of mutagens and carcinogens. CRC, Boca Raton, FL, pp 41–80

3. Miller J, Miller E (1977) Origins of human cancer. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 605–627

4. Miller EC, Miller JA (1981) Searches for ultimate chemical carcinogens and their reactions with cellular macromolecules. Cancer 47(10):2327–2345

5. Ames BN (1979) Identifying environmental chemicals causing mutations and cancer. Science 204(4393):587–593

6. OECD Test No. 487: In Vitro Mammalian Cell Micronucleus Test. OECD Publishing

7. Benfenati E (2013) Theory, guidance and applications on QSAR and REACH. ORCHESTRA consortium

8. Dearden JC, Barratt MD, Benigni R, Bristol DW, Combes RD, Cronin MT, Judson PN, Payne MP, Richard AM, Tichy M (1997) The development and validation of expert systems for predicting toxicity. Workshop (ECVAM Workshop 24). p 2

9. Benigni R, Bossa C, Tcheremenskaia O, Giuliani A (2010) Alternatives to the carcinogenicity bioassay: *in silico* methods, and the in vitro and in vivo mutagenicity assays. Expert Opin Drug Metab Toxicol 6(7):809–819. doi: 10.1517/17425255.2010.486400

10. Toropov AA, Toropova AP, Benfenati E (2009) Additive SMILES-based carcinogenicity models: probabilistic principles in the search for

robust predictions. Int J Mol Sci 10(7): 3106–3127

11. Rositsa S, Mojca FG, Andrew W (2010) Review of QSAR models and software tools for predicting genotoxicity and carcinogenicity. JRC scientific and technical reports-EUR 24427 EN. Publications Office of the European Union, Luxembourg

12. Benfenati E, Benigni R, Demarini DM, Helma C, Kirkland D, Martin TM, Mazzatorta P, Ouédraogo-Arras G, Richard AM, Schilter B, Schoonen WG, Snyder RD, Yang C (2009) Predictive models for carcinogenicity and mutagenicity: frameworks, state-of-the-art, and perspectives. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 27(2):57–90. doi:10.1080/10590500902885593

13. Ferrari T, Gini G, Benfenati E (2009) Support vector machines in the prediction of mutagenicity of chemical compounds. Proc NAFIPS 2009, June 14–17, Cincinnati, USA. pp 1–6

14. Ferrari T, Gini G, Bakhtyari NG, Benfenati E (2011) Mining toxicity structural alerts from SMILES: a new way to derive structure activity relationships. Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium. pp120–127.doi:10.1109/cidm.2011.5949444

15. Toxtree v, Ideaconsult Ltd, Sofia BAahtsn

16. Woo Y-T, Lai DY, Argus MF, Arcos JC (1995) Development of structure-activity relationship rules for predicting carcinogenic potential of chemicals. Toxicol Lett 79(1):219–228

17. OECD (2010) OECD Toolbox. http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm. Accessed 20 June 2015

18. Helma C (2006) Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. Mol Divers 10(2):147–158

19. Klopman G, Rosenkranz HS (1994) Approaches to SAR in carcinogenesis and

mutagenesis. Prediction of carcinogenicity/mutagenicity using MULTI-CASE. Mutat Res 305(1):33–46

20. Enslein K, Gombar VK, Blake BW (1994) Use of SAR in computer-assisted prediction of carcinogenicity and mutagenicity of chemicals by the TOPKAT program. Mutat Res 305(1):47–61

21. Smithing MP, Darvas F (1992) HazardExpert: an expert system for predicting chemical toxicity. ACS Symposium series American Chemical Society

22. Sanderson D, Earnshaw C (1991) Computer prediction of possible toxic action from chemical structure; the DEREK system. Hum Exp Toxicol 10(4):261–273

23. Ridings J, Barratt M, Cary R, Earnshaw C, Eggington C, Ellis M, Judson P, Langowski J, Marchant C, Payne M (1996) Computer prediction of possible toxic action from chemical structure: an update on the DEREK system. Toxicology 106(1):267–279

24. Helma C (2005) *In silico* predictive toxicology: the state-of-the-art and strategies to predict human health effects. Curr Opin Drug Discov Dev 8(1):27–31

25. Helma C, Cramer T, Kramer S, De Raedt L (2004) Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds. J Chem Inf Comput Sci 44(4):1402–1411. doi:10.1021/ci034254q

26. VEGA hwv-qe

27. Benigni R (2008) The Benigni/Bossa rulebase for mutagenicity and carcinogenicity—a module of Toxtree. JRC Scientific and Technical Reports, pp 1–78

28. ANTARES (2012) http://www.antares-life.eu/. Accessed 20 June 2015

29. Chemical carcinogens structures and experimental data (ISSCAN) (2008) ISS. http://www.epa.gov/ncct/dsstox/sdf_isscan_external.html. Accessed 20 June 2015

30. Kirkland D, Aardema M, Henderson L, Müller L (2005) Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: I. Sensitivity, specificity and relative predictivity. Mutat Res 584(1):1–256

31. Mayer J, Cheeseman MA, Twaroski ML (2008) Structure-activity relationship analysis tools: validation and applicability in predicting carcinogens. Regul Toxicol Pharmacol 50(1):50–58. doi:10.1016/j.yrtph.2007.09.005

32. EPA Environmental Protection Agency. http://www.epa.gov/nrmrl/std/qsar/qsar.html. Accessed 20 June 2015

33. IdeaConsult (2009) Toxtree user manual. IdeaConsult. https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/Toxtree_user_manual.pdf. Accessed 20 June 2015

34. The Carcinogenic Potency Database (CPDB) B, CA (USA). http://potency.berkeley.edu