ProtoICH

QPRF DOSSIER

www.protoqsar.com

Centro Europeo de Empresas Innovadoras (CEEI)
Parque Tecnológico de Valencia
Avda. Benjamín Franklin 12
46980 Paterna (Valencia, Spain)

info@protoqsar.com

+34 962 021 811

ProtoQSAR
Chemoinformatics and modeling solutions to optimize your chemicals

1. Substance

1.1. CAS:

109-66-0

1.2. EC number:

203-692-4

1.3. Chemical name:

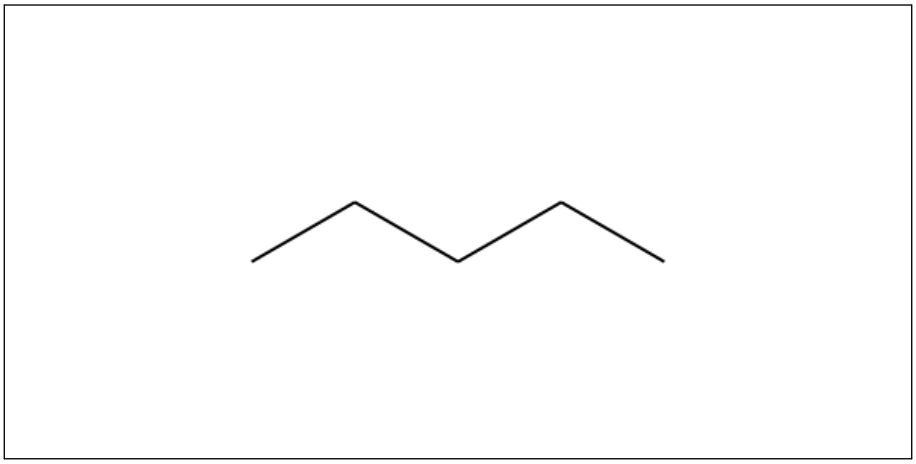
Pentane

1.4. Structural formula:

C5H12

1.5. Structure codes:

- SMILES: CCCCC



- Stereochemical features: There is no stereochemical information codified in the SMILES, so the substance is a non-stereochemical molecule or a racemic mixture

2. General information

2.1. Date of QPRF:

21-Mar-2022

2.2. QPRF autor and contact details:

- Autorship: ProtoQSAR

- Address: CEEI-Valencia Benjamin Franklin, 12.46980 Paterna

- Phone: +34 962 021 811

- Email: info@protoqsar.com

3. Prediction

3.1. Endpoint (OECD Principle 1):

- Endpoint: Mutagenicity (Ames test)

- Dependent variable: Binary classification: positive (mutagenic) / negative (non-mutagenic).

3.2. Algorithm (OECD Principle 2):

- Model or submodel name: ProtoQSAR model for Mutagenicity (Ames test)

- Model version: Version v2.0, calculated with ProtoQSAR proprietary software (ProtoPRED) (v4.0)

- Reference to QMRF: The corresponding QMRF, named ProtoQSAR model for Mutagenicity (Ames test) has been downloaded from https://protopred.protoqsar.com. More details can be requested to the owners of the model, ProtoQSAR 2000, S.L., by email info@protoqsar.com.

- Predicted value (model result): Non-mutagenic

- Predicted value (model comments): A compound is classified Ames positive if it significantly induces revertant colony growth in at least one of out of five strains tested following OECD Test N0. 471 guidance.

- Input for prediction: SMILES file containing structure of point 1.5

- Descriptors values: These are the calculated values of the descriptors of the model for the predicted molecule (after imputation and scaling)

Descriptor	Value
GATS5c	-0.38
GATS1d	2.17
BELd-1	-0.15
D/Dr3	-0.21
SIC1	-2.49
N-071	-0.22
MAXDN	-1.46
MATS1s	-1.89
BELc-1	2.85
BELi0	-3.04
MIC1	-2.18
AATS1i	-0.04
AATSC1i	-0.40

3.3. Applicability domain (OECD Principle 3):

- *Domain:* The applicability domain of the prediction was calculated by using the Kernel Density Estimation (KDE) method, the Tanimoto coefficient assessment, the Euclidean distance calculation and the Leverage analysis.
The compound falls inside the aplicability domain, at least by the Tanimoto

- *Structural analogues:* Similar structural analogues to the test compound have been found inside the training set. A table with the structures is shown in Annex 1

- *Considerations in structural analogues:* Whereas similar structural analogues have been found inside the training test, nor the predicted value neither the applicability domain of the model is not only calculated by structural analogy.

3.4. Uncertainty of the prediction (OECD Principle 4):

The models included in this study have been validated (both internal and external validation). Accuracy values are shown in the tables below (extracted from the QMRF document)

Parameters	Training	Validation
Accuracy (ACC)	0.92	0.77
Sensitivity, recall or true positive rate (TPR)	0.93	0.80
Specificity or true negative rate (TNR)	0.90	0.75
Precision or positive predictive value (PPV)	0.92	0.79
Negative predictive value (NPV)	0.92	0.76
Miss rate or false negative rate (FNR)	0.07	0.20
Fall-out or false positive rate (FPR)	0.10	0.25
False discovery rate (FDR)	0.08	0.21
False omission rate (FOR)	0.08	0.24
F-score	0.92	0.79
Matthews Correlation Coefficient (MCC)	0.84	0.55
Critical Success Index (CSI)	0.86	0.66
Area under the ROC (AUC)	0.92	0.77

Training set:

Experimental values	QSAR predictions		
	non-mutagenic	**mutagenic**	
non-mutagenic	2034	218	**90.3% (TNR)**
mutagenic	179	2438	**93.2% (TPR)**
Total (%)	91.9 (% pred NEG)	91.8 (% pred POS)	**91.8% (ACC)**

Validation set:

Experimental values	QSAR predictions		
	non-mutagenic	**mutagenic**	
non-mutagenic	562	187	**75.0% (TNR)**
mutagenic	179	696	**79.5% (TPR)**
Total (%)	75.8 (% pred NEG)	78.8 (% pred POS)	**77.5% (ACC)**

3.5. The chemical and biological mechanisms according to the model underpinning the predicted result (OECD Principle 5):

N/A

4. Adequacy

4.1. Regulatory purpose:

This prediction has been performed with the purpose of follow the guidelines for ICH M7 regulation.

4.2. Approach for regulatory interpretation of the model result:

According to the ICH-M7 regulation, if not experimental data on mutagenicity is available, an *in silico* approach combining two types of models: SAR models (based on expert rules) and QSAR models (statistical) can be used.

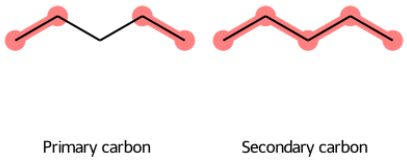
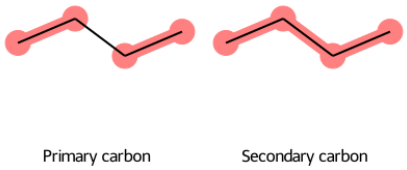
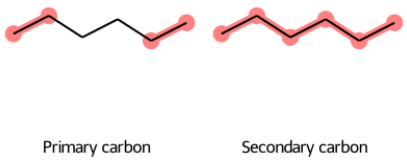
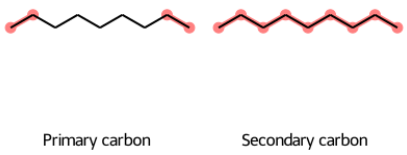
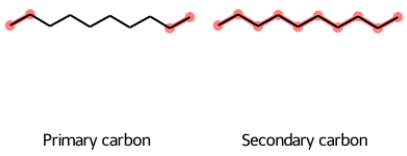
4.3. Outcome:

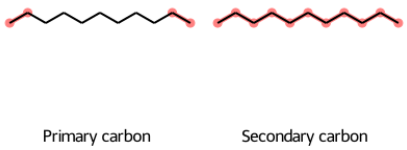
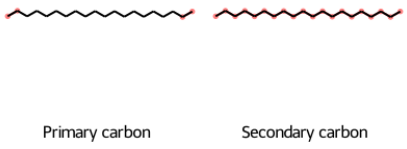
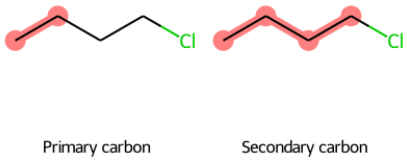
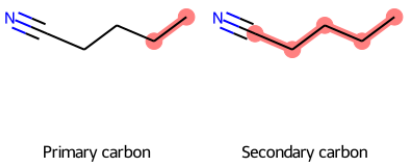
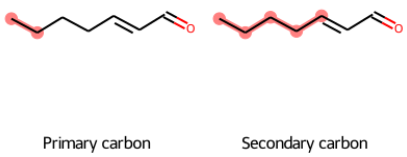
The ProtoQSAR model has been used for the prediction of Mutagenicity (Ames test) for CCCCC. The compound has been predicted as Non-mutagenic and the prediction falls inside the applicability domain.

4.4. Conclusion:

For ICH-M7 regulatory purposes, the negative prediction (non-mutagenic) reported above should be confirmed by a matching result from the expert rules based model. In case of contradiction, the result is considered uncertain

Annex 1: Structural Analogs

analog 1	SMILES:	CCCCC
		
	Observed value:	Negative
analog 2	SMILES:	CCCC
		
	Observed value:	Negative
analog 3	SMILES:	CCCCCC
		
	Observed value:	Negative
analog 4	SMILES:	CCCCCCCCC
		
	Observed value:	Negative
analog 5	SMILES:	CCCCCCCCCC
		
	Observed value:	Negative

analog 6	SMILES:	CCCCCCCCCCC
		
	Observed value:	Negative
analog 7	SMILES:	CCCCCCCCCCCCCCCCCCCC
		
	Observed value:	Negative
analog 8	SMILES:	CCCCCl
		
	Observed value:	Negative
analog 9	SMILES:	CCCCC#N
		
	Observed value:	Negative
analog 10	SMILES:	CCCC/C=C/C=O
		
	Observed value:	Positive