

ToxTwin

Rapport d'analyse toxicologique prédictive

Workflow SMILES → Scoring → Interprétation

Identifiant	Molécule	Date	Modèle
RPT-2026-001	Candidat inconnu	Mars 2026	ToxGNN-V1
Demandeur	Twingital Ventures	Statut	Confidentiel

1. Workflow d'analyse — De la structure au score

Ce document illustre le workflow complet d'utilisation de la plateforme ToxTwin pour l'analyse toxicologique prédictive d'un candidat médicament à partir de sa représentation SMILES. Il sert de modèle pour les analyses de routine.

1.1 Étape 1 — Génération et validation du SMILES

Le SMILES (Simplified Molecular Input Line Entry System) est la représentation textuelle standardisée de la structure chimique. Avant toute soumission à ToxTwin, la validité du SMILES doit être vérifiée via un outil de génération/vérification chimique.

Outils recommandés pour la génération/vérification SMILES

Outil	Type	Usage
ChemDraw (PerkinElmer)	Desktop · commercial	Dessin structure → SMILES · standard industrie pharma
MarvinSketch (ChemAxon)	Desktop + Web · freemium	Dessin structure → SMILES · vérification valence
RDKit (Python)	Bibliothèque open source	MolFromSmiles() → validation · canonicalisation · 2D/3D
PubChem Sketcher	Web · gratuit	pubchem.ncbi.nlm.nih.gov/edit3 · résolution DCI → SMILES
JSME / Ketcher	Web · open source	Éditeur embarquable dans applications web

Vérification RDKit locale

Avant soumission à ToxTwin, vérifier la validité du SMILES :

```
from rdkit import Chem
mol = Chem.MolFromSmiles("OCNC[F+]C1=CNCC1c2ccpcn2")
if mol: print("SMILES valide :", Chem.MolToSmiles(mol))
else: print("SMILES invalide")
```

Résultat de validation

SMILES valide · Canonicalisé : OCNC[F+]C1=CNCC1c1ccpcn1 · RDKit accepte [F+] et le cycle pyridine (p minuscule)

1.2 Étape 2 — Soumission à ToxTwin

Une fois le SMILES validé, la soumission à l'API ToxTwin s'effectue via une requête POST ou via l'interface Swagger UI.

Via curl (terminal)

```
curl -X POST http://localhost:8000/v1/score-toxicity \
  -H "Content-Type: application/json" \
  -d '{"query": "OCNC[F+]C1=CNCC1c2ccpcn2"}'
```

Via Swagger UI (navigateur)

- Ouvrir <http://localhost:8000/docs>
- Déplier POST /v1/score-toxicity → Try it out
- Saisir {"query": "OCNC[F+]C1=CNCC1c2ccpcn2"} → Execute

Via Python (intégration programmatique)

```
import requests
r = requests.post("http://localhost:8000/v1/score-toxicity",
                  json={"query": "OCNC[F+]C1=CNCC1c2ccpcn2"})
result = r.json()
```

2. Fiche moléculaire — Candidat RPT-2026-001

2.1 Structure et identification

Paramètre	Valeur
SMILES soumis	<chem>OCNC[F+]C1=CNCC1c2ccpcn2</chem>
SMILES canonique (RDKit)	<chem>OCNC[F+]C1=CNCC1c1ccpcn1</chem>
InChIKey	GVROQKCGIKPLDB-UHFFFAOYSA-N
Source de résolution	smiles_input (direct RDKit — aucun appel réseau)
CID PubChem	Non référencé (nouveau candidat)
MW calculée	Non disponible (structure inconnue des bases)
Temps d'inférence	582.8 ms (résolution directe SMILES, sans appel PubChem)

2.2 Caractéristiques structurales notables

Motif structural	Implication toxicologique potentielle
[F+] — fluor chargé positivement	Agent alkylant potentiel · réactivité avec l'ADN et les protéines · signal SR-p53 attendu
Cycle pipéridine (C1=CNCC1)	Amine secondaire cyclique · métabolisme oxydatif possible · formation de N-oxyde
Cycle pyridine (c2ccpcn2)	Hétérocycle azoté · faible réactivité propre · possible inhibiteur CYP450
Groupe hydroxyle (OC-)	Solubilité aqueuse favorisée · potentiel de conjugaison glucuronide
Amine (NHC-)	Basicité · absorption GI · risque d'interaction hERG si lipophile

2.3 Données expérimentales disponibles

Aucune donnée expérimentale dans la base Silver

Ce composé n'est pas référencé dans PubChem, ChEMBL ou Tox21. Toutes les prédictions ci-dessous sont des extrapolations du modèle ToxGNN-V1. Aucune LD50, aucun profil organique, aucun test Ames disponible. C'est précisément le cas d'usage principal de ToxTwin : scorer avant tout test expérimental.

3. Domaine applicatif — Validité de la prédiction

AVERTISSEMENT — Composé hors domaine applicatif

Similarité Tanimoto maximale : 0.153 (seuil : 0.30). Ce composé est structurellement très éloigné des 4 074 molécules du training set ToxGNN-V1. Les prédictions constituent des extrapolations et doivent être interprétées avec prudence. La validation expérimentale est obligatoire avant toute décision réglementaire.

La similarité Tanimoto de 0.153 signifie que la molécule la plus proche dans le training set ne partage que ~15% des sous-structures circulaires (Morgan FP, rayon 2) avec le candidat. Cette situation est fréquente pour les nouvelles entités chimiques (NCE) avec des groupes fonctionnels inhabituels comme [F+].

3.1 Cause de la sortie de domaine

Facteur	Impact sur le domaine applicatif
[F+] (fluor cationique)	Très rare dans les médicaments approuvés · sous-représenté dans Tox21/ChEMBL
Combinaison pipéridine + pyridine + F+	Pharmacophore inhabituel · pas de précédent structurel proche dans le training set
Absence dans PubChem/ChEMBL	Nouveau candidat propriétaire · aucun analogue connu référencé

3.2 Incertitude épistémique (Monte Carlo Dropout)

L'incertitude MC Dropout reflète la dispersion des 20 prédictions stochastiques. Sur ce composé hors domaine, les incertitudes les plus élevées correspondent aux endpoints les plus actifs — comportement attendu et cohérent.

Endpoint	Score	Incertitude	Niveau d'incertitude
SR-p53	32.6%	± 6.2%	Élevée (> 5%) — interpréter avec prudence
SR-ARE	25.5%	± 6.0%	Élevée (> 5%) — interpréter avec prudence
NR-AR	13.0%	± 7.1%	Élevée (> 5%) — interpréter avec prudence
SR-HSE	13.1%	± 3.5%	Modérée
Autres endpoints (8)	< 15%	< 3.5%	Acceptable

4. Résultats ToxGNN-V1 — Scoring multi-endpoints

Les 12 scores Tox21 prédits par ToxGNN-V1 sont présentés ci-dessous, classés par niveau de signal. Chaque score représente la probabilité que la molécule soit active sur l'assay correspondant.

4.1 Tableau de scoring complet

Endpoint	Score	Signal	Incertitude	Interprétation
SR-p53 (génotoxicité)	33%	●●●●○	± 6.2%	Modéré
SR-ARE (stress oxydatif)	25%	●●●●○	± 6.0%	Modéré
NR-AhR (récepteur Ah)	15%	●●●○○	± 3.2%	Faible
SR-HSE (choc thermique)	13%	●●●○○	± 3.5%	Faible
NR-AR (androgène)	13%	●●●○○	± 7.1%	Faible
SR-ATAD5 (réplication ADN)	10%	●●○○○	± 1.4%	Négligeable
NR-AR-LBD (androgène LBD)	8%	●●○○○	± 3.3%	Négligeable
NR-ER (œstrogène)	7%	●●○○○	± 1.8%	Négligeable
NR-Aromatase	6%	●●○○○	± 1.6%	Négligeable
NR-ER-LBD (œstrogène LBD)	4%	●○○○○	± 1.1%	Négligeable
SR-MMP (mitochondrie)	1%	●○○○○	± 0.2%	Négligeable
NR-PPAR-gamma	1%	●○○○○	± 0.4%	Négligeable

Légende	< 10%	10-25%	25-50%	> 50%
Niveau	Négligeable	Faible	Modéré — alerte	Fort — priorité

4.2 Analyse des signaux prioritaires

Signal 1 — SR-p53 : 32.6% (modéré · incertitude élevée 6.2%)

Le récepteur p53 est le gardien du génome. Son activation indique un stress génotoxique potentiel — l'ADN est endommagé ou la réplication est perturbée. Le groupement [F+] (fluor cationique) est un électrophile puissant, susceptible d'alkyler les bases de l'ADN ou les résidus cystéine des protéines. Ce signal est le plus préoccupant de la molécule et doit être investigué en priorité.

Recommandation expérimentale — SR-p53

Test Ames (mutagénicité bactérienne) · Test Comet (cassures ADN) · Micronoyaux in vitro · À réaliser avant tout passage in vivo

Signal 2 — SR-ARE : 25.5% (modéré · incertitude élevée 6.0%)

L'Antioxydant Response Element (ARE) régule les gènes de défense contre le stress oxydatif (Nrf2/KEAP1). Son activation à 25.5% indique que la molécule peut induire un stress oxydatif ou être reconnue comme électrophile par la machinerie cellulaire. Ce signal est cohérent avec la présence de [F+] et complète le tableau génotoxique.

Recommandation expérimentale — SR-ARE

Test ROS (espèces réactives de l'oxygène) · Glutathion cellulaire (GSH) · Activité NQO1 · Test DPPH si applicable

Signaux secondaires (10-20%) — NR-AhR, SR-HSE, NR-AR

NR-AhR (15.4%) : le récepteur aryl hydrocarbène est sensible aux composés aromatiques plans. Le cycle pyridine peut contribuer à ce signal, mais il reste en zone faible. SR-HSE (13.1%) : la voie du choc thermique indique un stress protéique modéré — non prioritaire. NR-AR (13.0% · incertitude 7.1%) : signal androgène faible mais avec forte incertitude — à surveiller sans être alarmant.

Signaux non significatifs (< 10%)

Les 7 endpoints restants (SR-ATAD5, NR-AR-LBD, NR-ER, NR-Aromatase, NR-ER-LBD, SR-MMP, NR-PPAR-gamma) présentent des scores < 10% avec des incertitudes faibles. Ces signaux sont considérés comme non significatifs dans le contexte de cette analyse préliminaire.

5. Synthèse et recommandations

5.1 Profil toxicologique prédictif

Dimension	Niveau	Commentaire
Génotoxicité (SR-p53)	MODÉRÉ	Signal [F+] · investigation obligatoire avant Phase I
Stress oxydatif (SR-ARE)	MODÉRÉ	Cohérent avec [F+] électrophile · tester GSH/ROS
Perturbation endocrinienne	FAIBLE	NR-AhR et NR-AR en zone faible · pas d'alerte majeure
Toxicité mitochondriale	NÉGLIGEABLE	SR-MMP 1.4% · pas de signal
Données expérimentales	ABSENTES	Nouveau candidat · aucune LD50, aucun Ames disponible
Domaine applicatif	HORS DOMAINE	Tanimoto 0.153 · extrapolation · prudence requise
Incertitude globale	MODÉRÉE	Mean ± 3.0% · acceptable mais > seuil 5% sur SR-p53/ARE

5.2 Plan d'expérimentation recommandé

Sur la base du profil prédictif ToxGNN-V1, le plan d'études expérimentales suivant est recommandé avant toute décision de développement :

Prio.	Étude	Signal adressé	Rationalité
P1	Test Ames (OCDE 471)	SR-p53 : 32.6%	Mutagenicité bactérienne · [F+] potentiellement alkylant
P1	Test Comet (OCDE 489)	SR-p53 : 32.6%	Cassures ADN simples et doubles brins · in vitro
P1	Test ROS / GSH	SR-ARE : 25.5%	Stress oxydatif · consommation glutathion cellulaire
P2	Micronoyaux in vitro	SR-p53 + SR-ARE	Aberrations chromosomiques · confirme génotoxicité
P2	LD50 rat (OCDE 420)	Données manquantes	Toxicité aiguë · alimentera le Silver ToxTwin
P3	Inhibition hERG (patch-clamp)	Amine + cycle azoté	Cardiotoxicité potentielle · non prédit mais à vérifier
P3	Panel CYP450	Pyridine + amine	Interactions métaboliques · DDI potentielles

5.3 Conclusion

Verdict ToxTwin v1.0 — Candidat RPT-2026-001

Profil de risque MODÉRÉ sur deux voies génotoxiques (SR-p53 32.6%, SR-ARE 25.5%), cohérentes avec la présence du groupement [F+]. Les autres endpoints sont dans la zone faible à

négligeable. IMPORTANT : ce composé est hors du domaine applicatif du modèle (Tanimoto 0.153) — les prédictions sont des extrapolations. La progression vers les études réglementaires nécessite impérativement la validation expérimentale des signaux génotoxiques (Ames + Comet) avant toute décision de développement.

Note méthodologique : ce rapport a été généré automatiquement par ToxTwin v1.0 (ToxGNN-V1, Mean AUC Tox21 = 0.857) en 582 ms. Il constitue une aide à la décision et ne se substitue pas au jugement d'un toxicologue qualifié ni aux études réglementaires obligatoires (ICH S2, S7A, S7B).