



Université HASSAN II

Faculté des Sciences et Techniques
Mohammedia



Etude des Relations Quantitatives Structure Activité des Urées Cycliques, Cyanoguanidines Cycliques et Sulfamides Cycliques (Inhibiteurs VIH)

R. Mouhibi J. Rakik and M. Zahouily

Laboratoire de Catalyse, Chimométrie et Environnement



Introduction

QSAR

Résultat

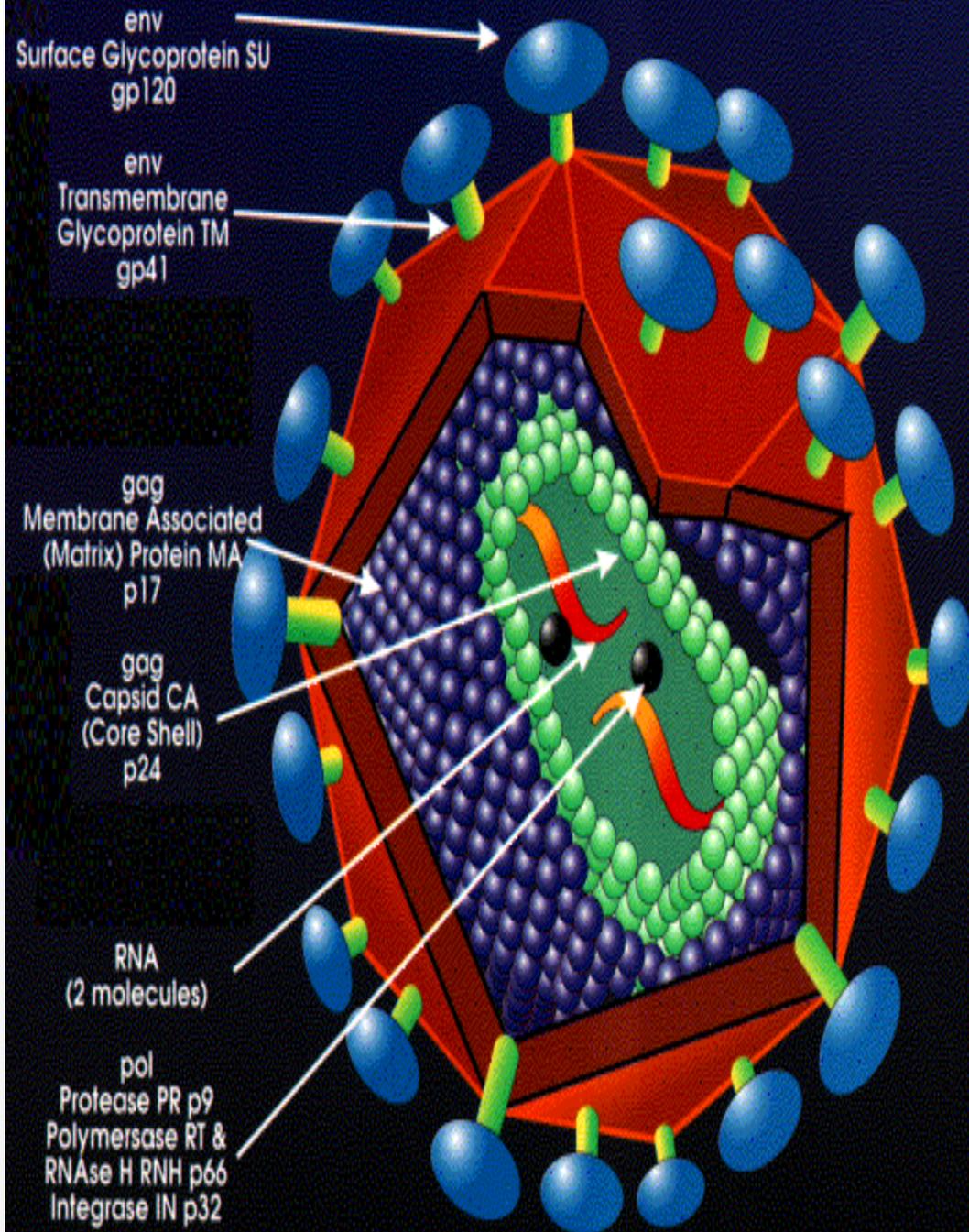
Conclusion



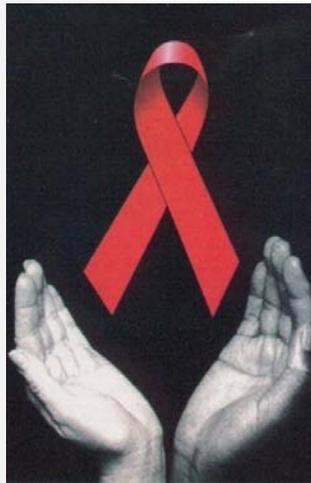
La découverte

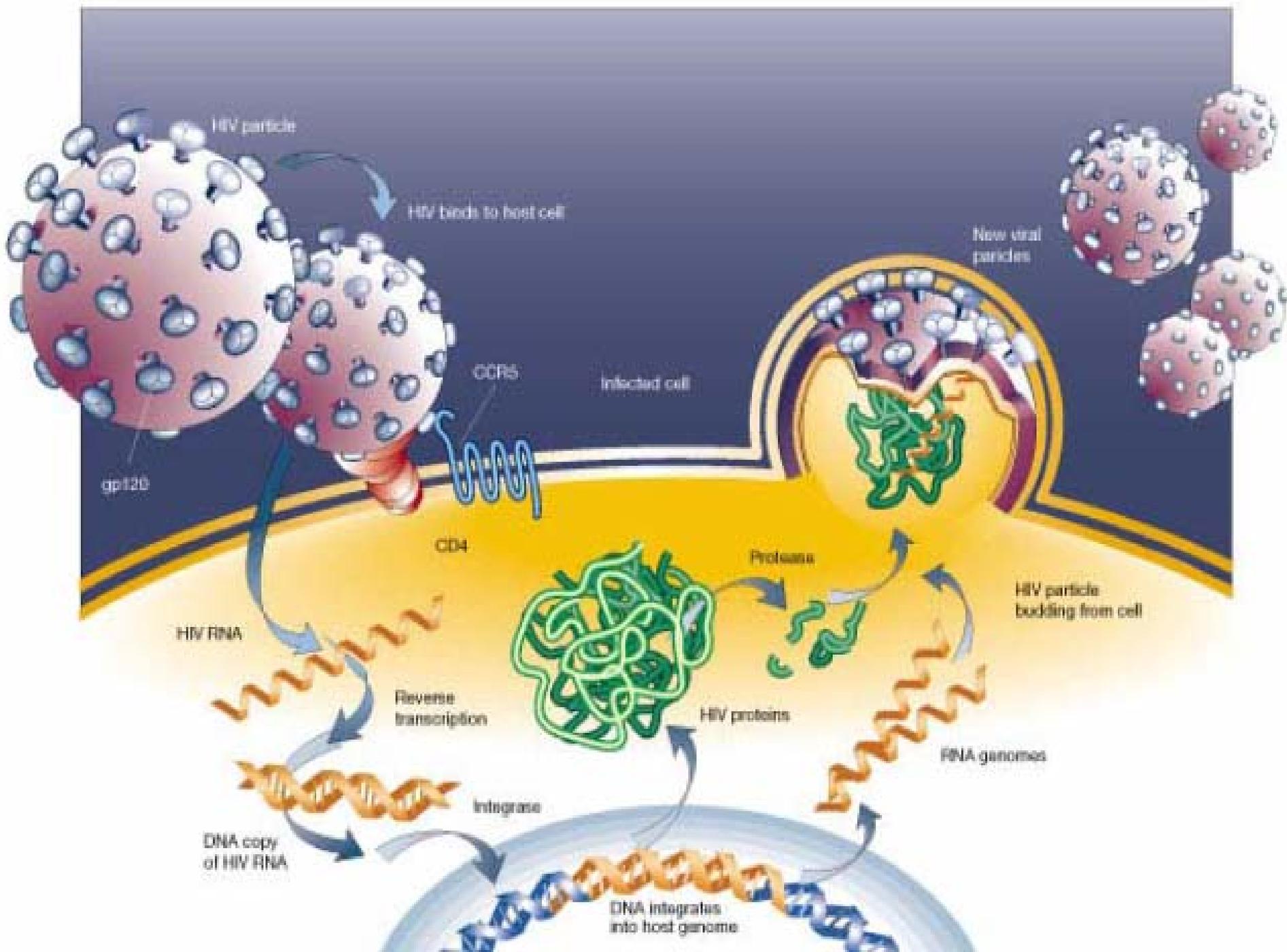
Virus: 1983

VIH: 1986



Cycle du Virus de l'Immunodéficience Humain (VIH)



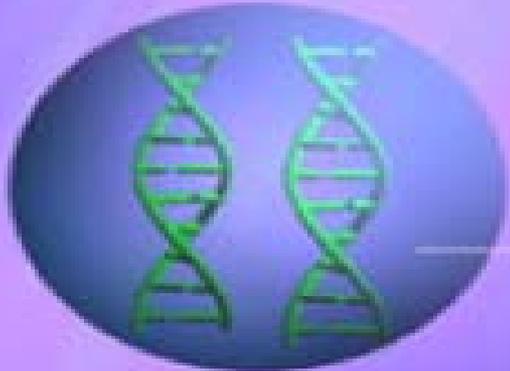




HIV

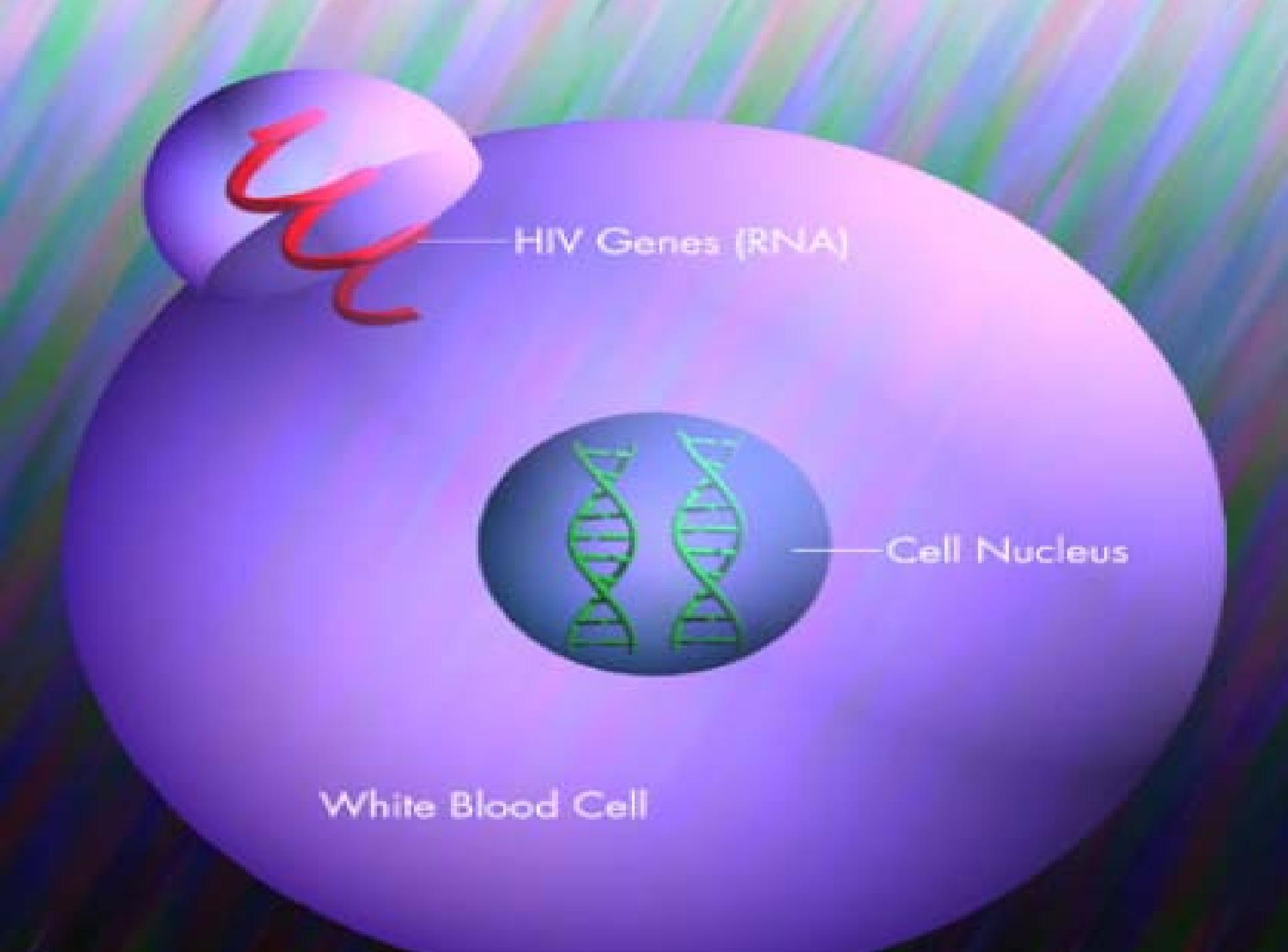


Cell Receptor



Cell Nucleus

White Blood Cell



HIV Genes (RNA)

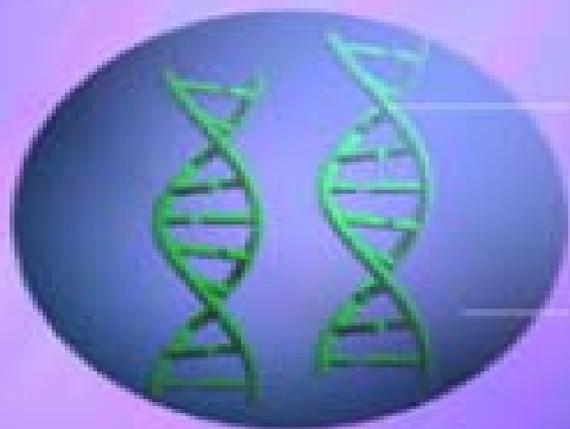
Cell Nucleus

White Blood Cell

White Blood Cell



HIV Genes (RNA)



Cell Genes (DNA)

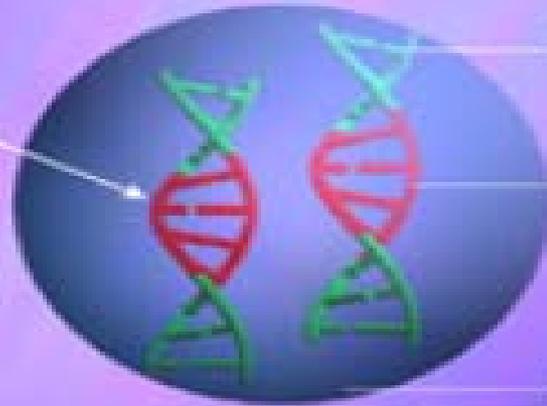
Cell Nucleus



HIV Genes (DNA)



HIV Genes (DNA)



Cell Genes (DNA)

HIV Genes (DNA)

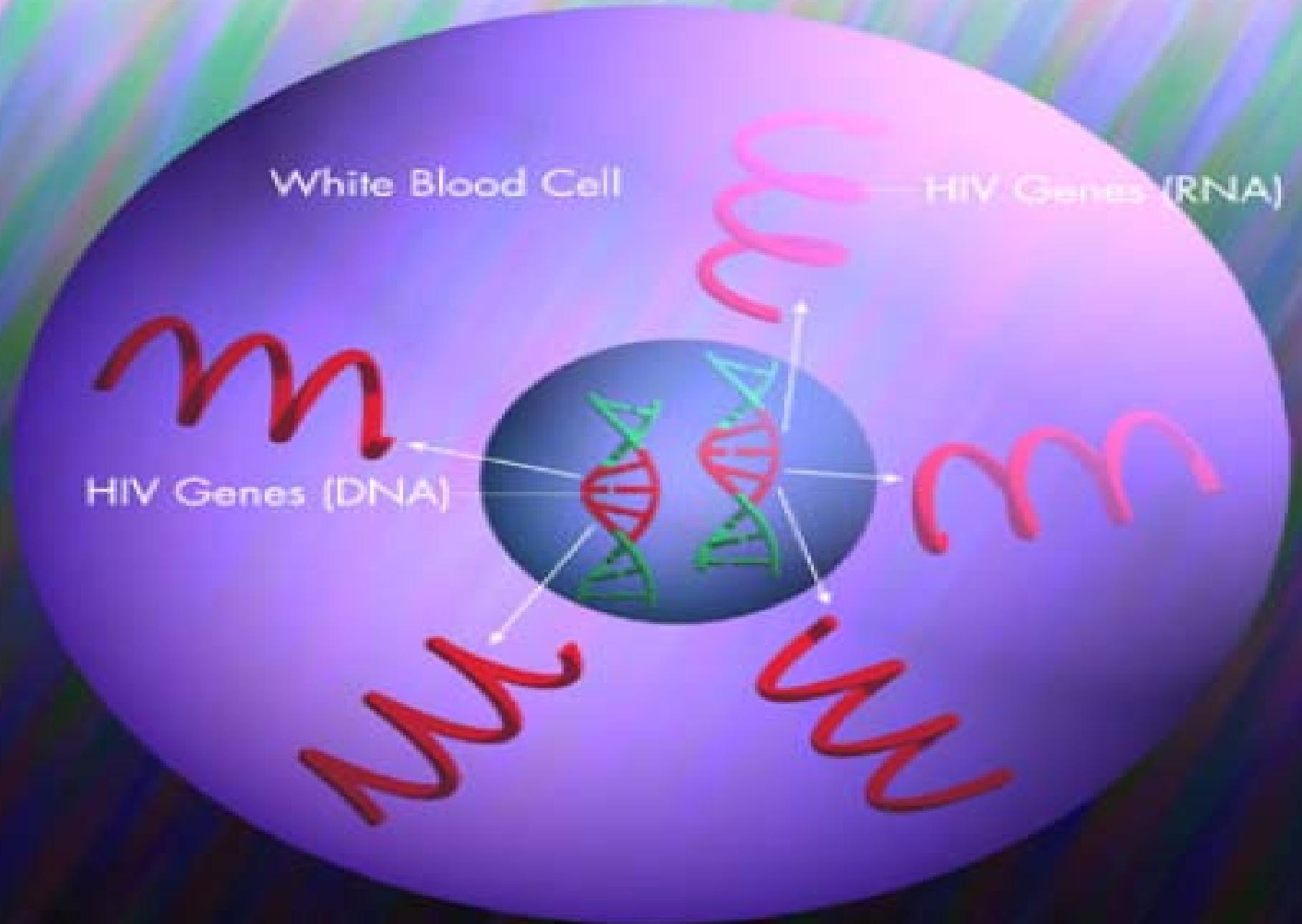
Cell Nucleus

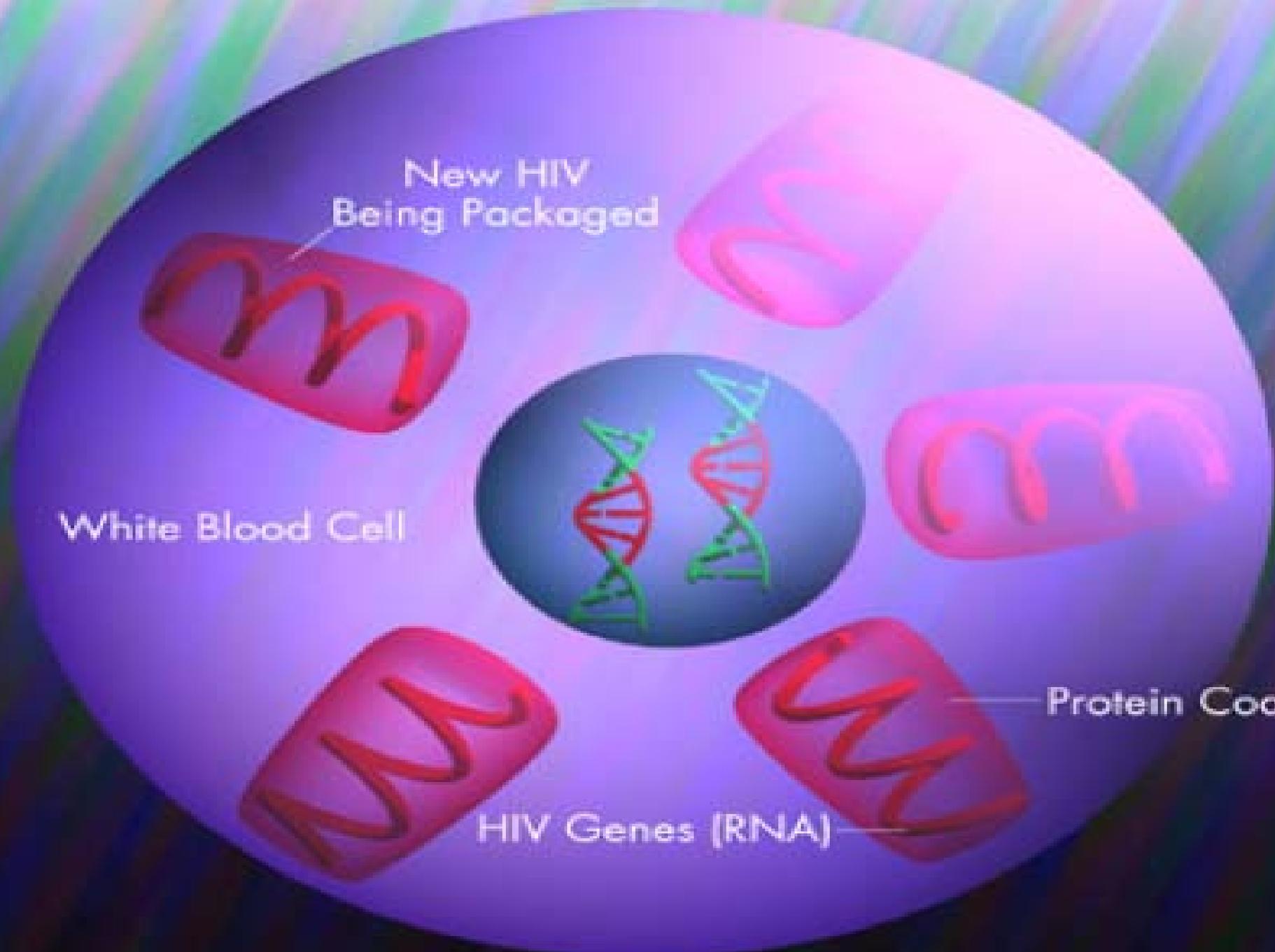
White Blood Cell

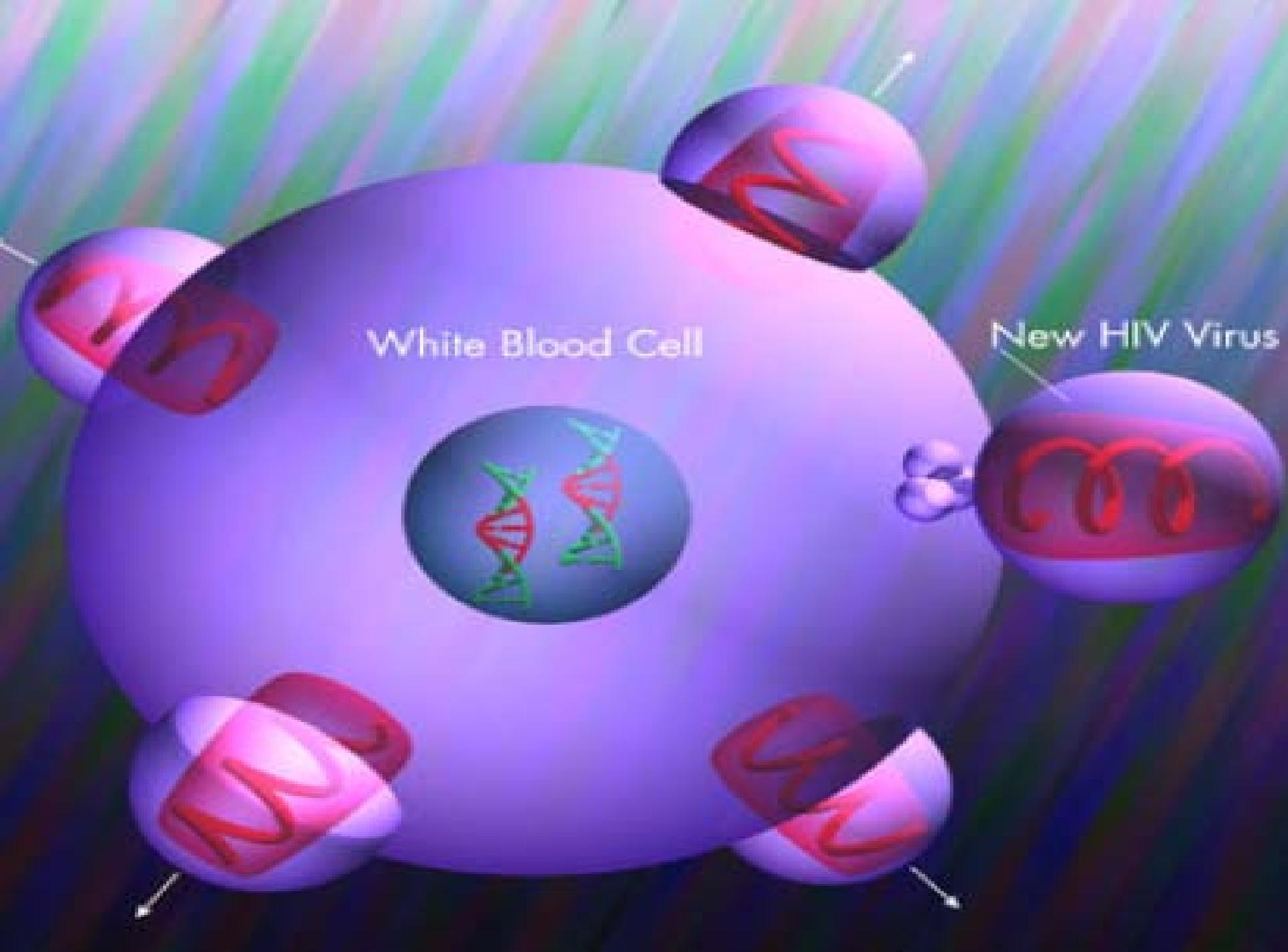
White Blood Cell

HIV Genes (RNA)

HIV Genes (DNA)







White Blood Cell

New HIV Virus



C'est quoi la protéase ?

- Enzyme brisant les liaisons peptidiques des protéines.
- Provoque la maturation des virions



Modélisation moléculaire

QSAR



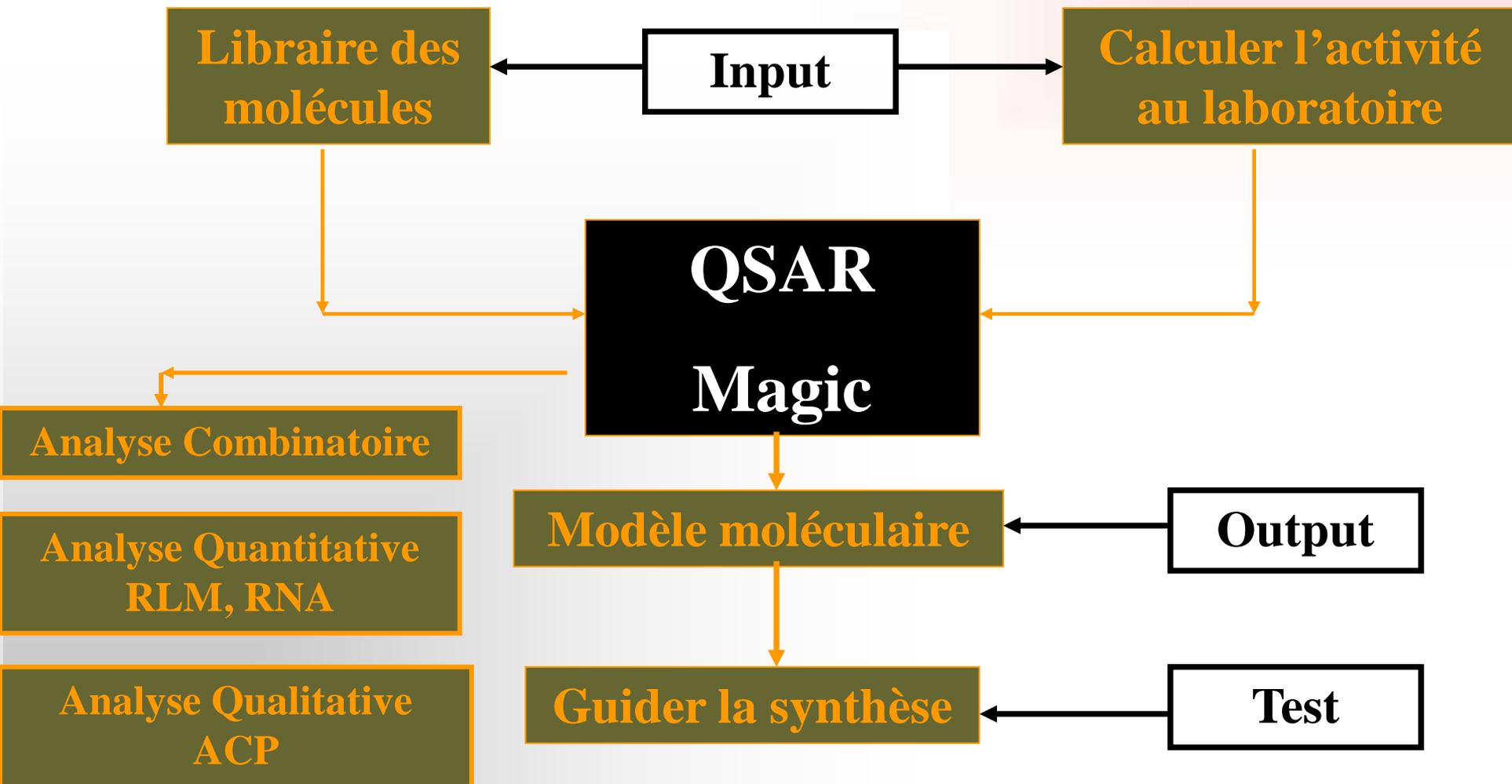
Relations Quantitatives Structure *Activité*

activité = f(propriétés physico-chimiques et/ou structurales)

- Comprendre la relation entre la structure et l'activité,
- Faire le design de composés avec une meilleure activité et prédire l'activité des composés avant leur synthèse.



Relations Quantitatives Structure - Activité





RESULTATS



Etude de la relation structure-activité

Echantillon

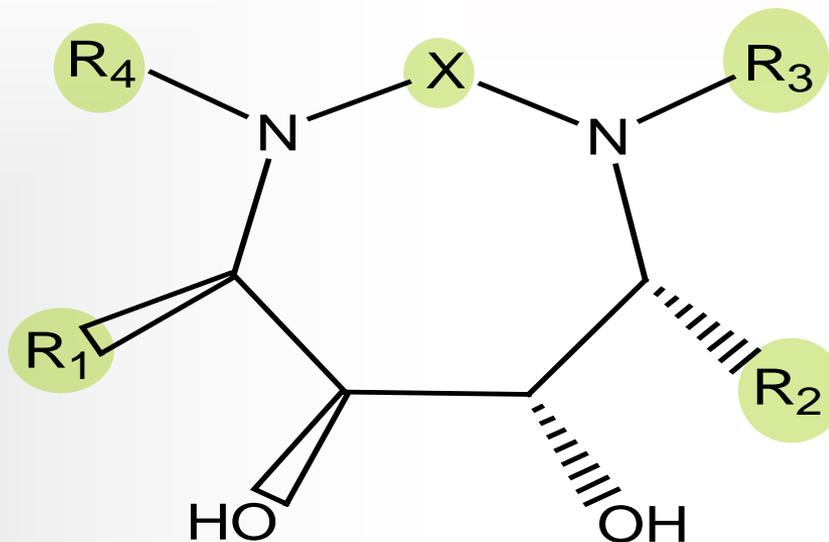
Activité

Descripteurs

Méthodes
d'analyse

➡ Taille : 160

➡ Squelette de base



R¹, R², R³, R⁴ et X sont des substituants variables



Etude de la relation structure-activité

Echantillon



Activité

Log($10^6/K_i$) variable dépendante

K_i : la concentration en inhibiteur pour laquelle la moitié des sites enzymatiques sont occupés

Activité



Gamme des activités

2,026



9

Descripteurs



Moyenne

5,49

**Méthodes
d'analyse**



Etude de la relation structure-activité

Echantillon

Activité

Descripteurs

**Méthodes
d'analyse**

Description locale



R₁, R₂, R₃, R₄ et X

- ➔ **Stériques (S, V, PM, ...)**
- ➔ **Lipophile (logP(R_i))**
- ➔ **Électronique (HOMO, LUMO...)**



Étude de la relation structure-activité

Échantillon

Activité

Descripteurs

**Méthodes
d'analyse**

Régression linéaire multiple (RLM)

Réseau de neurones artificiels (RNA)

Early-Stopping



Régression linéaire multiple

✚ Sélection par la régression pas à pas "stepwise"

logP(R₃), logP(X), DLH(R₂) et DLH

Lipophilie : logP(R₃) et logP(X)

Effet électronique : DLH(R₂) et DLH



Régression linéaire multiple

$$\begin{aligned} \text{Log}(10^6/\text{ki}) = & (5,29 \pm 0,058) + (0,219 \pm 0,07) * \text{DLH}(R_2) - \\ & (0,258 \pm 0,07) * \text{LogP}(R_3) + (1,149 \pm 0,077) * \text{logP}(X) \\ & + (0,488 \pm 0,074) * \text{DLH} \end{aligned}$$

$$n = 160 \quad r = 0,92 \quad s = 0,74193$$



Amélioration du modèle linéaire

Élimination de toutes les molécules ayant une erreur d_i définie comme suit :

$$D_i > 2s/\text{moy (obs)}$$

Avec

$$d_i = | \text{Obs}_i - \text{Cal}_i | / \text{Obs}_i$$



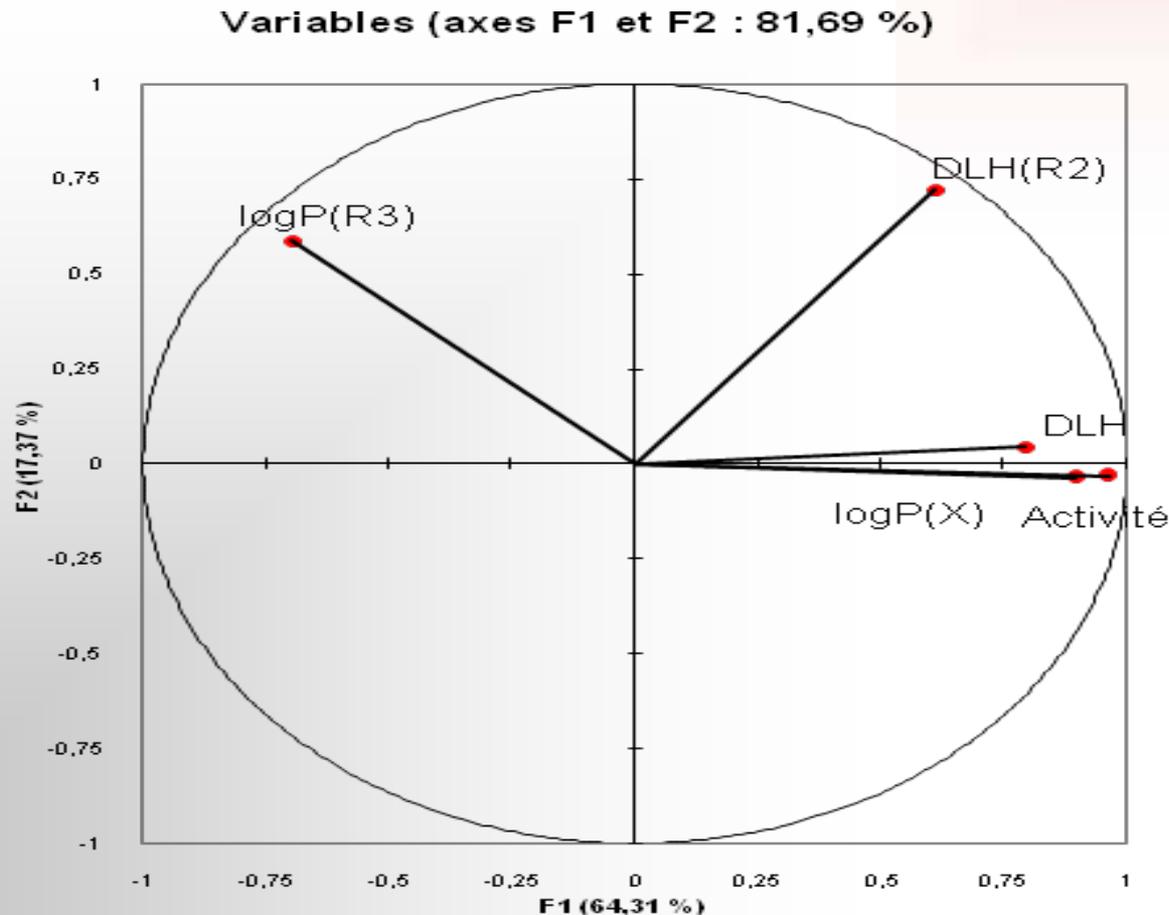
Amélioration du modèle linéaire

$$\begin{aligned} \text{Log}(10^6/\text{ki}) = & (5,49 \pm 0,047) + (0,165 \pm 0,057) * \text{DLH}(\text{R}_2) - \\ & (0,246 \pm 0,066) * \text{logP}(\text{R}_3) + \\ & (1,25 \pm 0,066) * \text{logP}(\text{X}) + (0,434 \pm 0,061) * \text{DLH} \end{aligned}$$

$$n = 142 \quad r = 0,954 \quad s = 0,65626 \quad q^2L_{20\%O} = 0,88$$



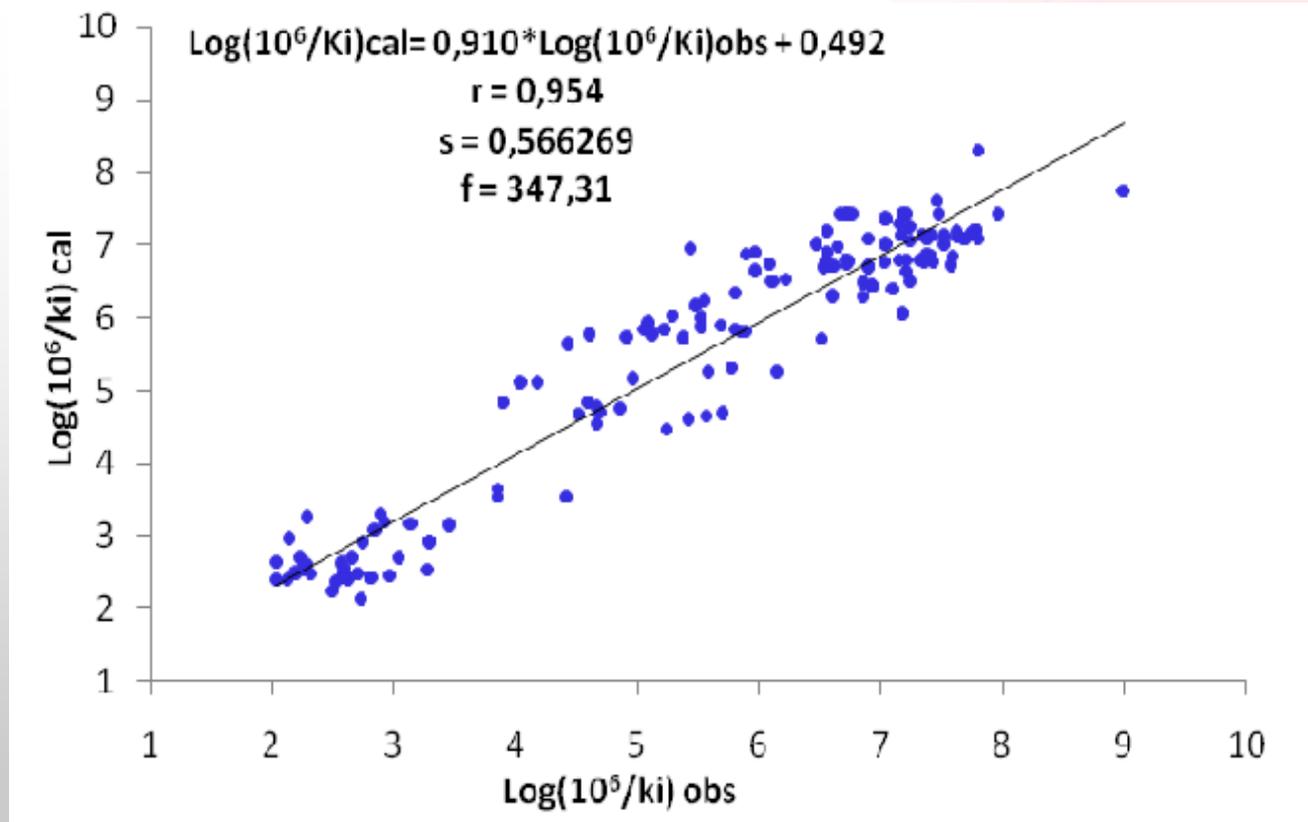
Analyse en composantes principales





Régression linéaire multiple

Appréciation de la corrélation



Bonne corrélation entre les valeurs calculées et les valeurs observées



Régression linéaire multiple

Calcul des contributions

Méthode de Gore $C_{\%} = a_i \cdot sd_i / \sum_i a_i \cdot sd_i$

Contributions des descripteurs pertinents

Descripteurs	Contribution
DLH(R ₂)	7 %
logP(R ₃)	11,05%
logP(X)	<u>62,22%</u>
DLH	19,73%

logP(X) > DLH > logP(R3) > DLH(R2)



Régression linéaire multiple

Validation du modèle linéaire

Méthode : Cross-validation leave-20%-out

$$q^2 = 1 - \left(\frac{PRESS}{Var} \right)$$

$q_{L20\%}^2 = 0,88 > 0,6$ (Wold) \rightarrow Bon modèle prédictif



Régression linéaire multiple

Test de crédibilité



$n = 142$ $r = 0,954$ $s = 0,566$

$n = 142$ $r = 0,5$ $s = 1,57$



Qualité statistique ↗

➤ Descripteurs retenus expliquent bien l'activité *anti-VIH* .



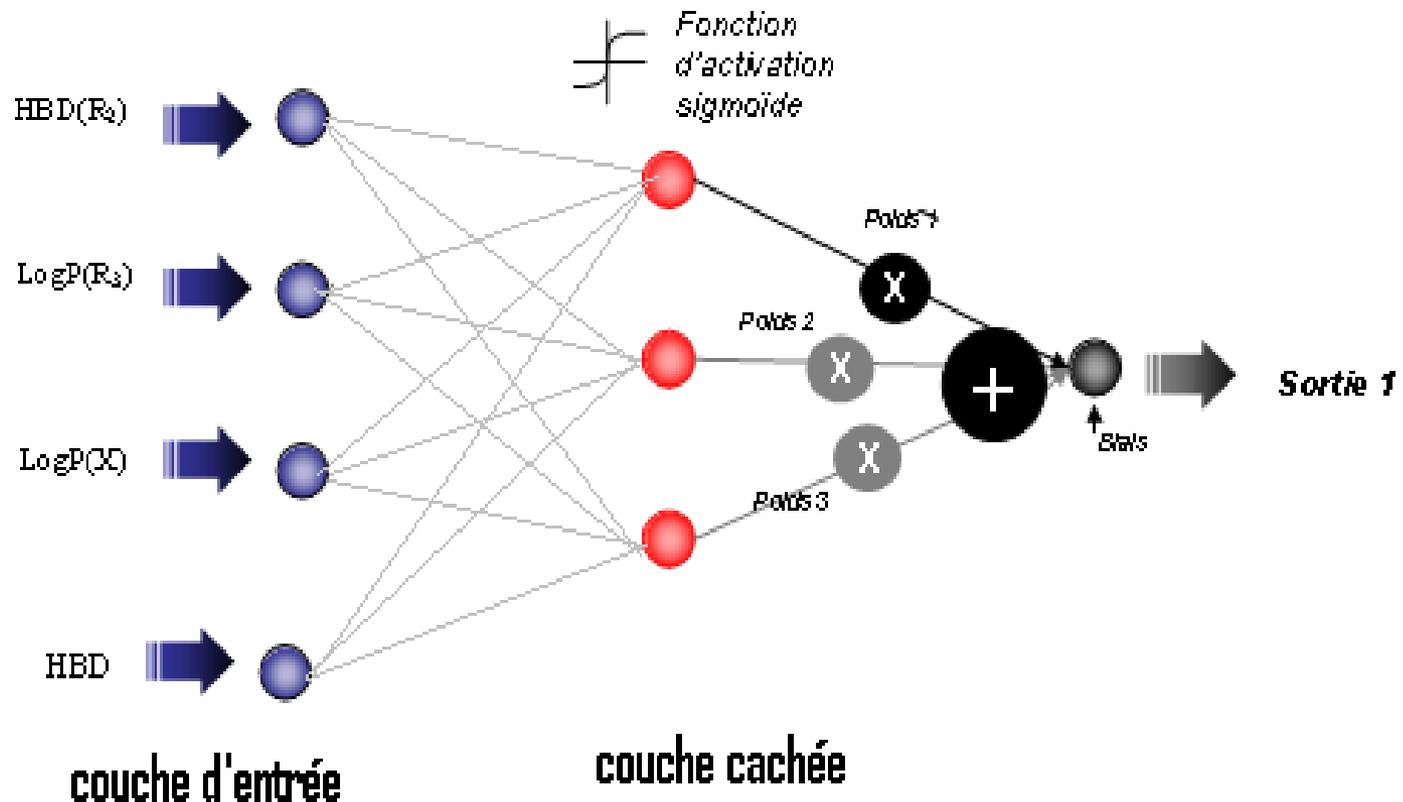
Analyse neuronale



Réseaux de neurones artificiels

Phase d'entraînement

Architecture [4-3-1] employée





Réseaux de neurones artificiels

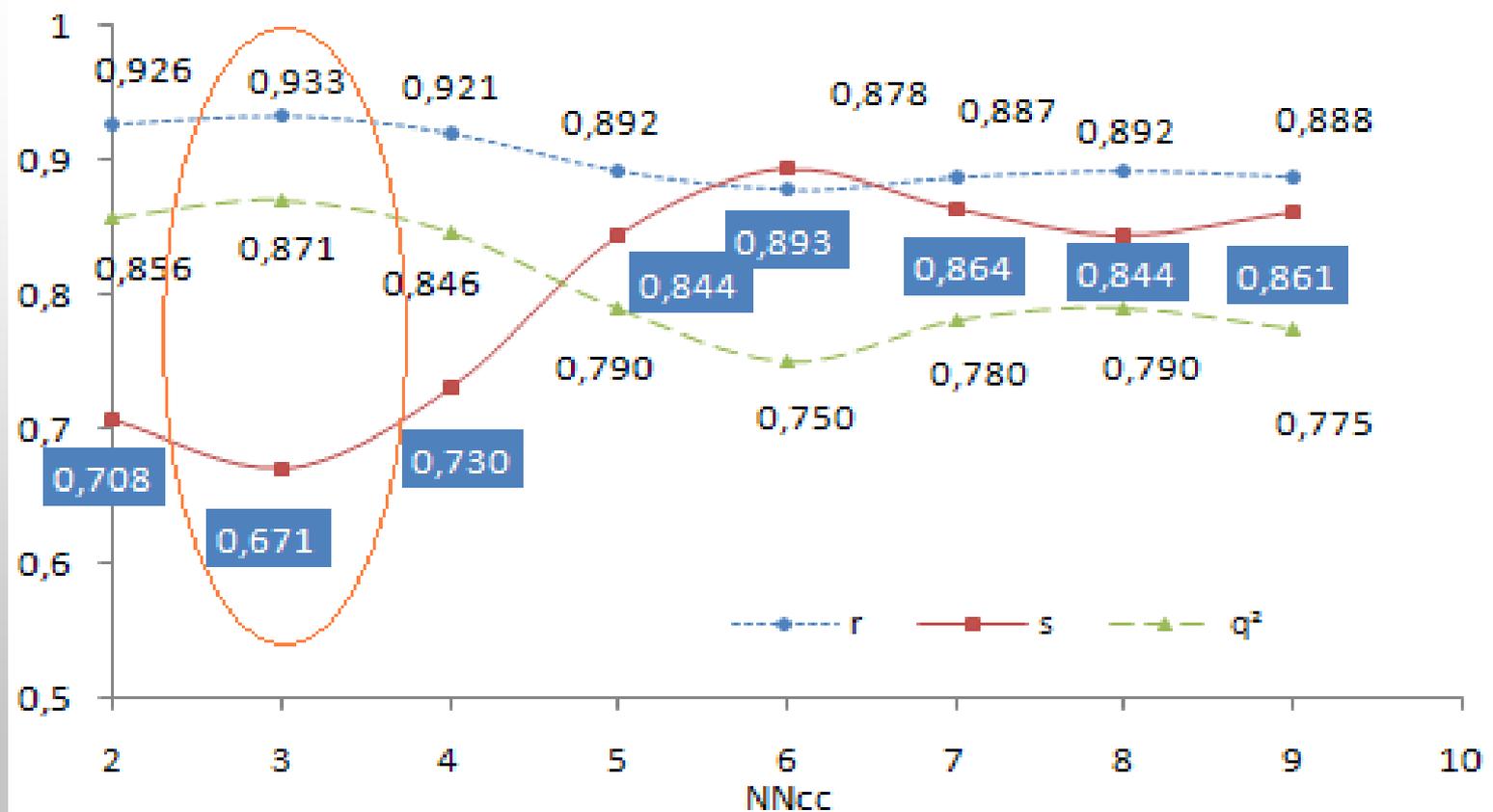
Arrêt Prématuré "Early-Stopping"

- Ensemble d'apprentissage : 71 molécules.
- Ensemble test : 36 molécules.
- Ensemble de validation : 35 molécules.



Réseaux de neurones artificiels

Régularisation Bayésienne





Réseaux de neurones artificiels

Phase d'entraînement

$r = 0,945$ et $s = 0,566$ RLM (91%)
 $r = 0,974$ et $s = 0,42$ RNA (95%)

Amélioration par rapport au modèle linéaire.

Existence éventuelle des relations non-linéaire entre l'activité et les descripteurs pertinents.



Réseaux de neurones artificiels

La comparaison de la RLM et le RNA

$\text{LogP}(X) > \text{DLH} > \text{LogP}(R_3) > \text{DLH}(R_2)$ RLM

$\text{LogP}(X) > > \Sigma [\text{DLH} + \text{DLH}(R_2)] > \text{LogP}(R_3)$ RNA



Changement de classification.

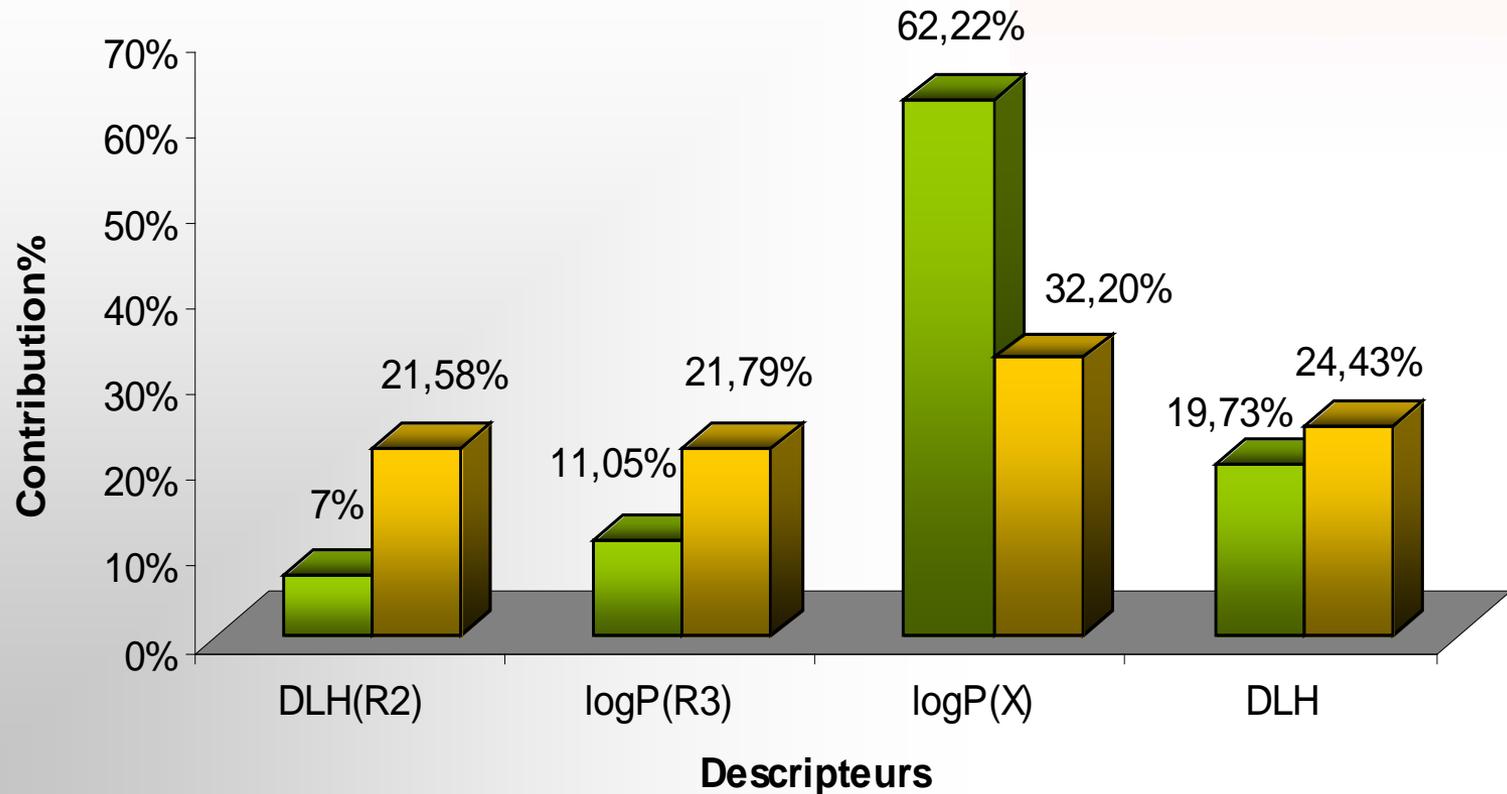


Relation non linéaire entre l'effet électronique et l'activité.



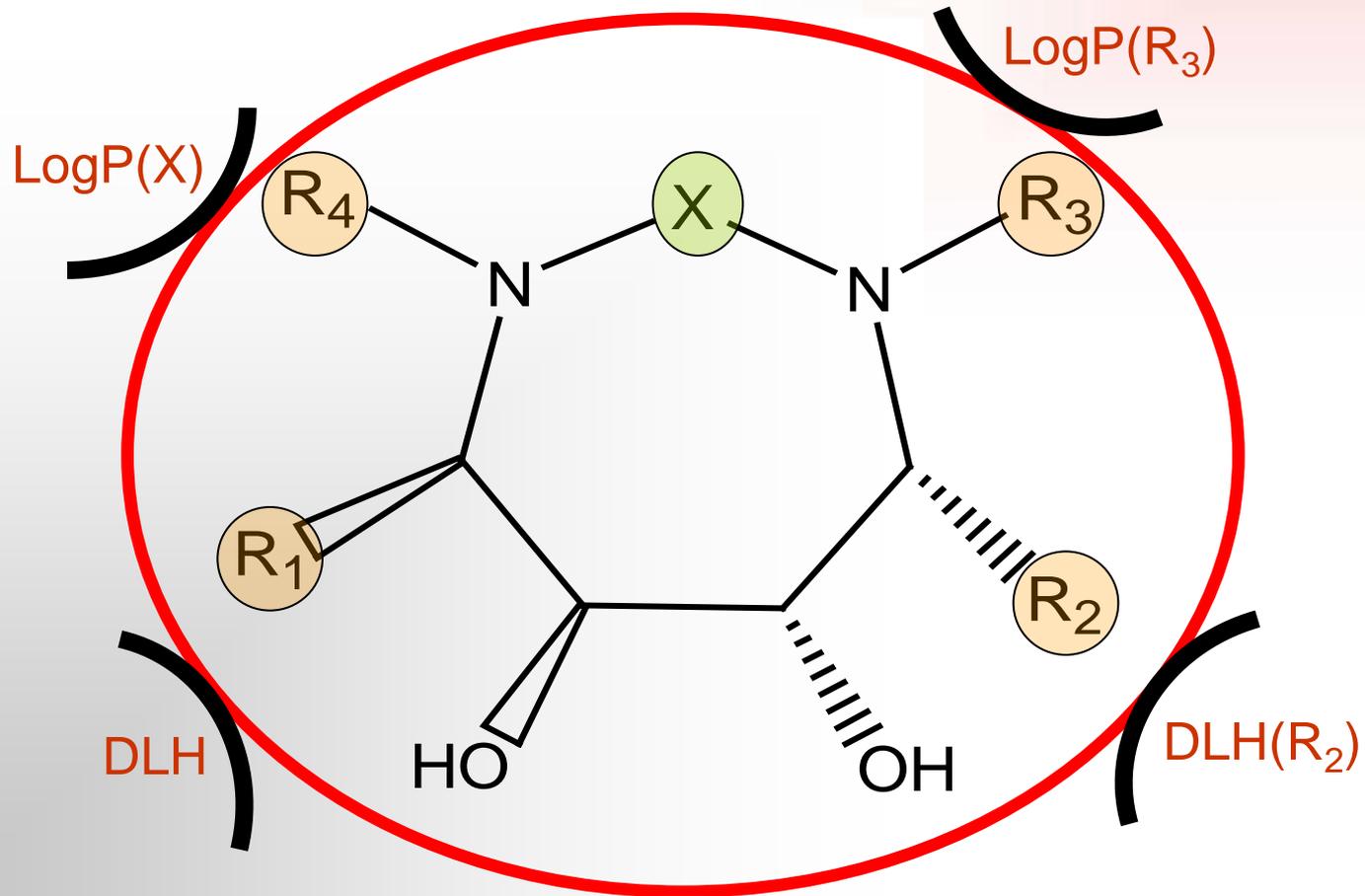
Contributions des descripteurs

La comparaison de la RLM et le RNA





Modèle structure activité anti-VIH





Conclusion

- + Descripteurs retenus expliquent bien l'activité inhibitrice.**
- + Les résultats du modèle linéaire sont confirmés et améliorés par l'application du réseaux de neurones.**
- + La régression linéaire multiple et les réseaux de Neurones artificiels sont deux méthodes d'analyses complémentaires**



Merci pour votre
attention



Merci pour votre
attention

