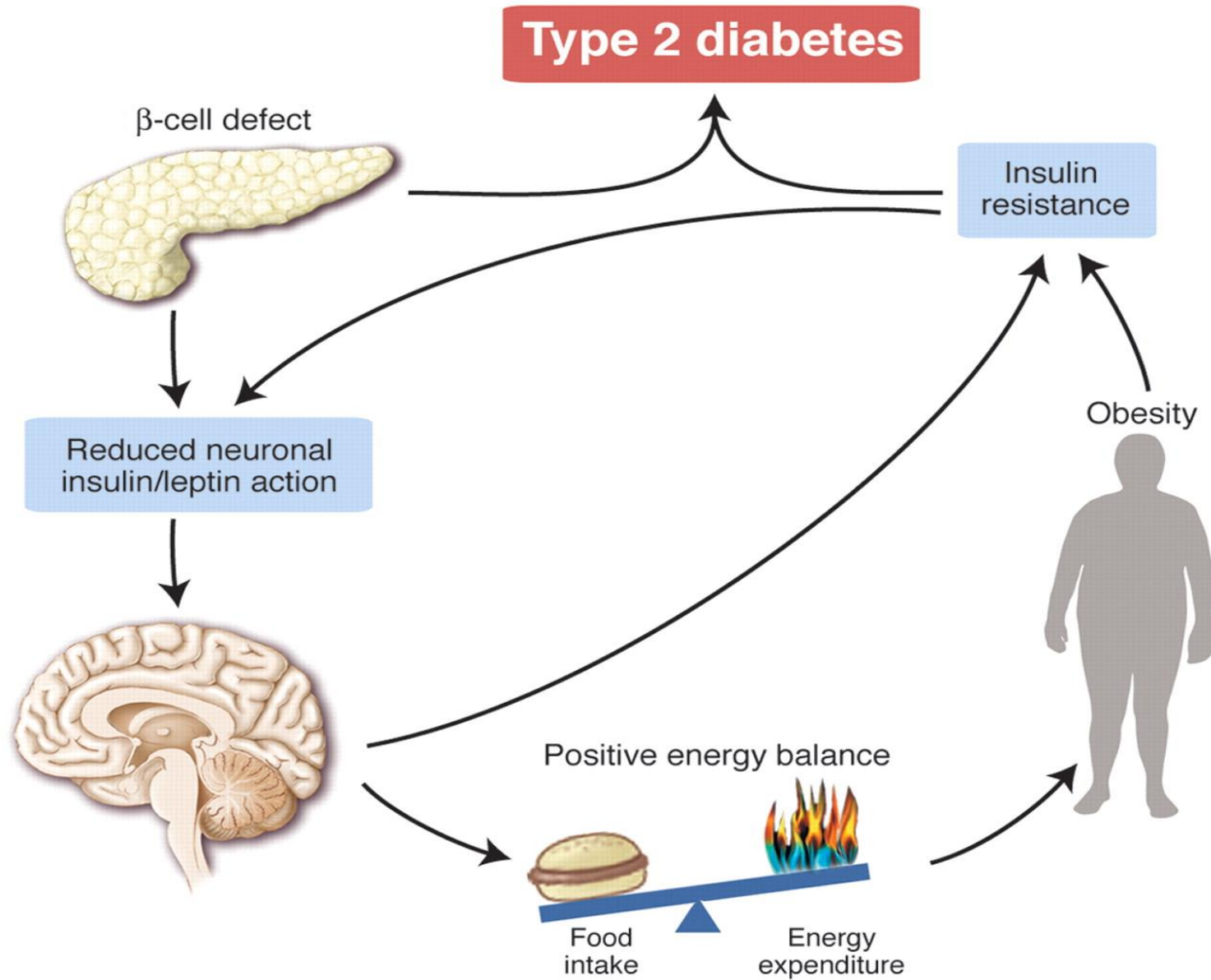


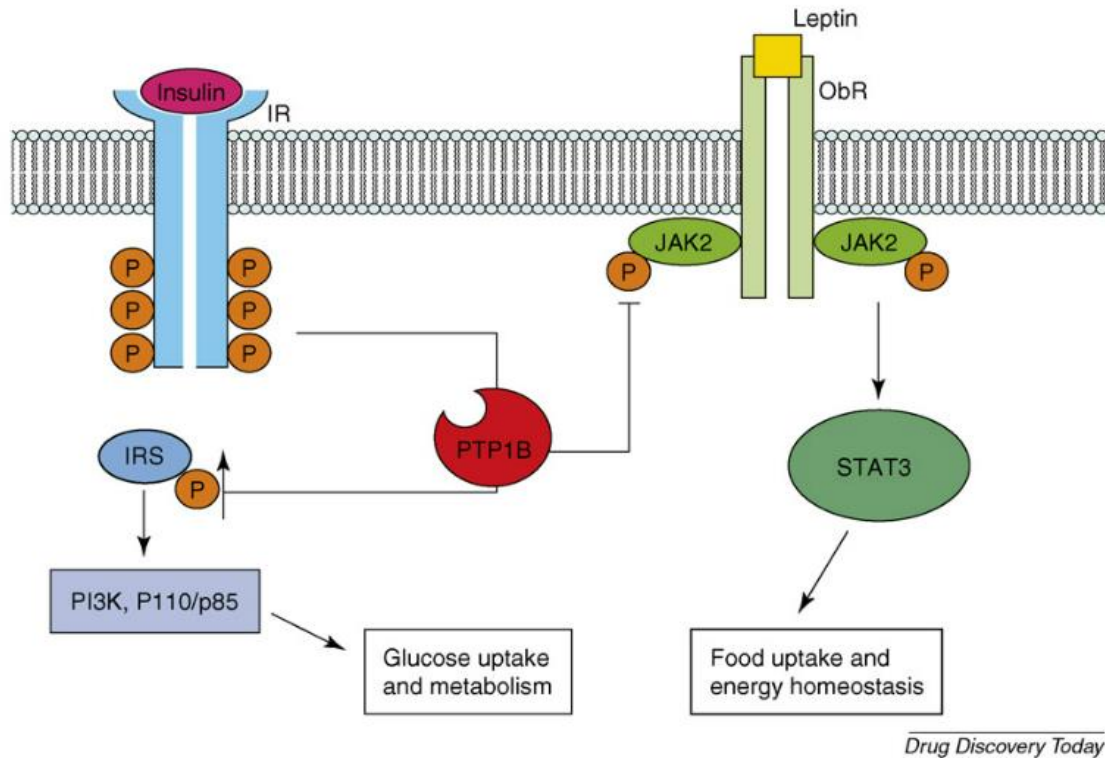
고급화학정보학

숭실대학교 생명정보학과 석사과정 박정운

Energy homeostasis



PTP-1B, a candidate for the treatment of obesity and diabetes



Sheng Zhang et al., *Drug discovery Today*, 2007

FIGURE 1

The role of PTP1B in insulin and leptin signalling. In the insulin signalling pathway, PTP1B can associate with and dephosphorylate activated insulin receptor (IR) or insulin receptor substrates (IRS) [26–30]. In the leptin pathway, PTP1B binds and dephosphorylates JAK2, which is downstream of the leptin receptor, ObR [40,41]. STAT3 and P110/p85 are downstream targets of JAK2 and IRS1, respectively. P110/p85 is a specific form of PI3K responsive to insulin signalling.


그렇다면 PTP-1B inhibitor과 IC_{50} 값 간의 relationship을 나타내는 model을 만들면 치료제를 개발하는 데에 도움이 되지 않을까?

Reference paper

ORIGINAL ARTICLE



Quantitative structure-activity relationship (QSAR) and design of novel ligands that demonstrate high potency and target selectivity as protein tyrosine phosphatase 1B (PTP 1B) inhibitors as an effective strategy used to model anti-diabetic agents

David Ebuka Arthur^{a,b} , Stephen Ejeh^c and Adamu Uzairu^c

^aDepartment of Chemistry, ABU zaria, Baze University, Abuja, Nigeria; ^bBaze University, Ahmadu Bello University Zaria, Abuja, Nigeria; ^cAhmadu Bello University, Zaria, Nigeria

ABSTRACT

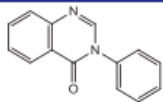
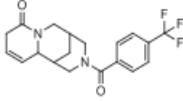
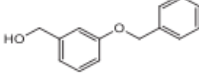
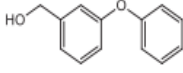
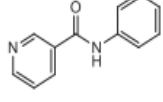
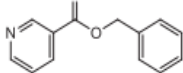
Diabetes and obesity have increased dramatically in recent decades worldwide. Diabetes mainly emerged as a major health care burden disease in both the US and other industrialized countries, among which type II diabetes is the most common. Discovering new and effective treatments for diabetes is currently a high international health priority. In the present study a computational technique was used to model 97 compounds with PTP-1B inhibitory activity, in order to demonstrate the Quantitative structure-activity relationship (QSAR) of these compounds a genetic function approximation (GFA) algorithm was applied to pick the best descriptors and multiple linear regression (MLR) was used to establish a relationship between the PTP-1B inhibitory activity of these compounds and the best molecular descriptors. This QSAR study allowed investigating the influence of very simple and easy-to-compute descriptors in determining biological activities, which shed light on the key factors that aid in the design of novel potent molecules using computer-aided drug design tools.

ARTICLE HISTORY

Received 17 December 2019
Accepted 17 April 2020

KEYWORDS

QSAR; ligand-based drug design; binding energy; molecular docking; molecular descriptors

| S/No | Pubchem CID | SID | Chemical Structure | IC ₅₀ [μM] | pIC ₅₀ |
|------|-------------|-----------|---|-----------------------|-------------------|
| 1* | 785275 | 170484936 |  | 692 | 3.159894 |
| 2 | 66563668 | 170484943 |  | 5453 | 2.263365 |
| 3 | 74341 | 170484942 |  | 6247 | 2.204328 |
| 4 | 26295 | 170484941 |  | 520 | 3.283997 |
| 5 | 74463 | 170484940 |  | 3471 | 2.459545 |
| 6 | 7191 | 170484939 |  | 818 | 3.087247 |

Study Design

1. Data collection :

- Reference paper에서 제공한 **PTP-1B inhibitor 97개의 분자**를 사용한다.
- 해당 데이터의 SID number를 이용해 **PubChem database**에서 3D SDF 파일을 다운로드 받아서 사용한다.

2. Molecular descriptor calculation :

- **PaDEL descriptor generator**를 이용해 2D descriptor 계산한다.
- SID number를 이용해 3D SDF 파일을 구했으므로, standardization 및 MM2 force field 과정은 생략한다.
- End point 값은 Reference paper에서 제공한 **pIC₅₀ 값**을 사용한다.

3. QSAR model development :

- **pyQSAR module**을 이용해서 jupyter notebook으로 진행한다. (python3)
- Reference paper에서 제시한 식과 비교하기 위해서 **10개의 descriptor**를 사용한다.
- Model을 생성할 때 학습 횟수를 1000번으로 진행하였고, k-fold cross validation 과정에서는 k = 5로 설정하여 진행한다. (나머지 parameter는 pyQSAR tutorial과 동일하게 설정)

Study Design

4. QSAR model 비교 :

- Reference paper에서 제시한 QSAR model과 pyQSAR를 이용해서 만든 **QSAR model을 비교**한다.
- Reference paper에서 제시한 threshold 값으로 model의 성능을 비교한다.

5. Outlier 정의 및 molecular 구조 확인 :

- pyQSAR에 의해 생성된 QSAR model을 이용해 예측 값과 End point 값의 차이를 구한 뒤 제곱을 한다. (Square of errors)
- **Square of error의 값이 3보다 큰 값들을 outlier로 정의한다.**
- Rdkit module을 이용해 **outlier molecular들이 공통적으로 가지고 있는 구조를 확인**한다.

6. QSAR remodeling

- Outlier를 제거한 데이터를 만든다.
- 해당 데이터를 이용해서 QSAR model를 만들고, Reference paper에서 제시한 식과 비교한다.

Result 1. Comparison between each QSAR model

$$\begin{aligned} \text{(a) } \text{pIC}_{50} = & -20.51146 + 3.24397(\mathbf{AATS0e}) \\ & + 0.00031(\mathbf{ATSC4v}) + 1.16318(\mathbf{ATSC1e}) \\ & + 3.30587(\mathbf{SpMin7_{Bhm}}) - 3.39864(\mathbf{SpMax7_{Bhe}}) \\ & + 6.37754(\mathbf{SpMin5_{Bhe}}) + 0.45604(\mathbf{nHBint8}) \\ & + 0.83362(\mathbf{naaNH}) - 14.29726(\mathbf{ETA_{ShapeX}}) \\ & - 3.78441(\mathbf{nT12Ring}) \end{aligned}$$

$$\begin{aligned} \text{(b) } \text{pIC}_{50} = & 0.45134884 - 1.7222851(\mathbf{ATSC3c}) \\ & + 3.17791184(\mathbf{ETA_dEpsilon_A}) + 7.69710481(\mathbf{SpMin5_Bhv}) \\ & + 1.71030924(\mathbf{VR2_DzZ}) + 1.65891677(\mathbf{VR2_Dzp}) \\ & + 2.1555729(\mathbf{maxHBint3}) - 2.54643557(\mathbf{nHBint4}) \\ & - 5.2441163(\mathbf{nT8Ring}) + 4.74055064(\mathbf{nTRing}) \\ & - 4.12663419(\mathbf{piPC5}) \end{aligned}$$

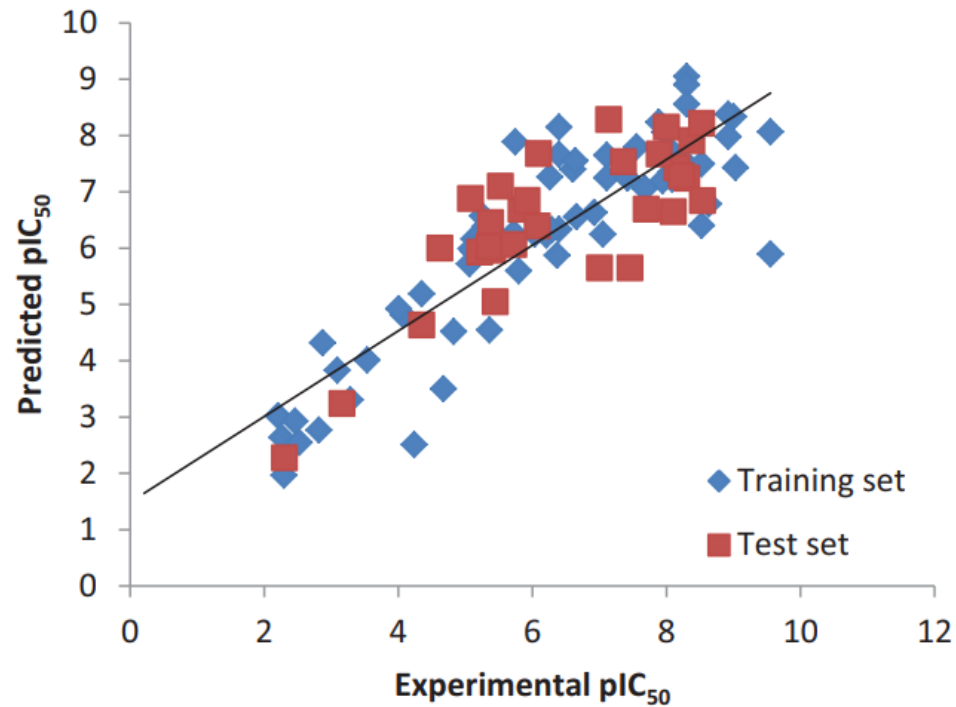
Figure 1 | The equations presented by each model

(a)는 Reference paper에서 제시된 model 식이다.

(b)는 pyQSAR module에서 제시된 model 식이다.

Result 1. Comparison between each QSAR model

(a)



(b)

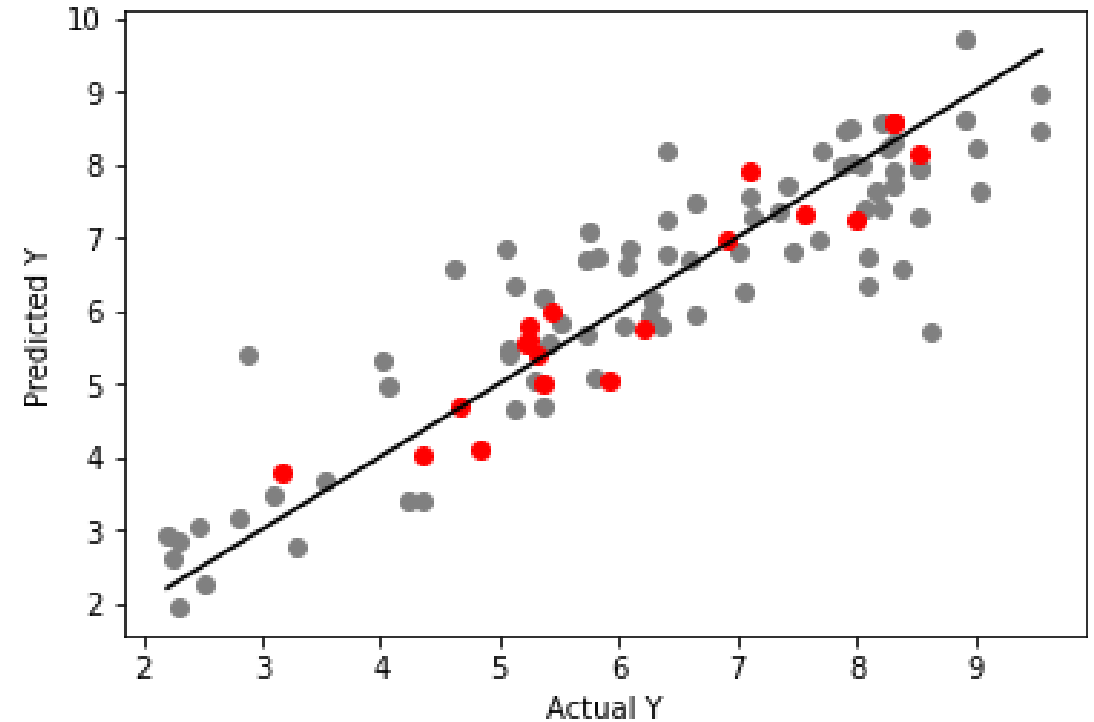


Figure 2 | Graph of End point values vs. Predicted values for each models

(a)는 Reference paper에서 제시된 model이며 x축은 End point 값, y축은 예측 값이다. (파란색은 Training set, 빨간색은 Test set이다.)

(b)는 pyQSAR module에서 제시된 model이며 x축은 End point 값, y축은 예측 값이다. (회색점은 Training set, 빨간색은 Test set이다.)

Result 1. Comparison between each QSAR model

(a) $R^2 = 0.819111$ RMSE CV : 0.877221

$Q^2 = 0.801812$

$R^2 - Q^2 = 0.017299$

(b) **Table 2.** GFA-linear regression results of the developed model, threshold values and sources.

| Symbol | Name | Values for the developed model | Threshold | Source |
|--------------|--|--------------------------------|-------------|--------|
| R^2 | Coefficient of determination | 0.7604 | ≥ 0.6 | a |
| Q^2 | LOO cross validation coefficient | 0.6653 | > 0.5 | b |
| R^2_{pred} | External test set's coefficient of determination | 0.6248 | ≥ 0.6 | a |
| $R^2 - Q^2$ | Difference between R^2 and Q^2 | 0.1251 | ≤ 0.3 | a |
| F value | Variation ratio | 101.0065 | High | a |
| K' | Intercept | 0.9901 | ≤ 1.15 | b |

a = (Veerasamy et al., [10]), b = (Golbraikh & Tropsha, [26]).

Figure 3 | Comparison of the performance of each model.

(a)는 pyQSAR module에서 제시된 model의 R^2 값, Q^2 값, R^2 과 Q^2 간의 차이 값이다.

(b)는 Reference paper에서 제시된 model의 R^2 값, Q^2 값, R^2 과 Q^2 간의 차이 값이다.

Result 3. Comparison between QSAR remodeling and reference

(a)

$$R^2 = 0.89811 \qquad RMSE \text{ CV} : 0.658561$$
$$Q^2 = 0.881059$$
$$R^2 - Q^2 = 0.017051$$

(b) **Table 2.** GFA-linear regression results of the developed model, threshold values and sources.

| Symbol | Name | Values for the developed model | Threshold | Source |
|--------------|--|--------------------------------|-------------|--------|
| R^2 | Coefficient of determination | 0.7604 | ≥ 0.6 | a |
| Q^2 | LOO cross validation coefficient | 0.6653 | > 0.5 | b |
| R^2_{pred} | External test set's coefficient of determination | 0.6248 | ≥ 0.6 | a |
| $R^2 - Q^2$ | Difference between R^2 and Q^2 | 0.1251 | ≤ 0.3 | a |
| F value | Variation ratio | 101.0065 | High | a |
| K' | Intercept | 0.9901 | ≤ 1.15 | b |

a = (Veerasamy et al., [10]), b = (Golbraikh & Tropsha, [26]).

Figure 5 | Comparison of the performance of each model.

(a)는 outlier를 제거한 데이터를 가지고 만든 model의 R^2 값, Q^2 값, R^2 과 Q^2 간의 차이 값이다.

(b)는 Reference paper에서 제시된 model의 R^2 값, Q^2 값, R^2 과 Q^2 간의 차이 값이다.

Discussion

1. 한계점 :

- pyQSAR module에서는 **사전에 데이터를 training set과 test set을 나눌 수가 없었다.** 이를 대체하기 위해서 k-fold cross validation 과정에서 training set과 validation set으로 나누어서 진행을 했다. (여기서 validation set은 test set으로 가리킨다고 봐도 무방하다.)
- Reference paper에서 제공한 데이터는 PumChem CID number도 존재하였는데, 총 97개의 데이터 중에 일부는 중복된 CID number를 가지고 있었으며 또한 **PumChem에서 제공되지 않은 CID number도 존재**하였다.
- PumChem에서 SID number를 가지고 SDF 파일을 다운받을 때, **2D SDF 파일을 제공하지 않아** Reference paper와 동일한 조건으로 진행 할 수 없었다. (Reference paper에서는 2D SDF 파일을 가지고 MM+ force field를 적용한 후 descriptor를 계산하였다.)

2. 추가 연구 :

- pyQSAR module을 통해 생성된 QSAR model이 Reference paper보다 성능이 더 좋다고 보여주었지만, **PTP-1B inhibitor들이 실제로 어떻게 PTP-1B에 결합을 하는지** 보여줄 필요성이 있다. (Validation 과정이 필요하다.)
- 따라서 Outlier molecule들이 어떻게 PTP-1B와 결합하는 지를 확인할 필요가 있으며, 이를 하기 위해서 **Molecular docking analysis**를 진행해야 할 필요가 있다.

Conclusion

1. **Diabetes과 Obesity는 Energy homeostasis가 깨져 일어나는 질병이다.**
2. 많은 연구자들이 해당 질병에 대한 치료제를 개발했지만, Insulin resistance 및 Leptin resistance에 의해서 치료 효과가 높지는 않았다.
3. **PTP1B**는 Insulin 및 Leptin signaling pathway를 조절하며 이를 통해 glucose metabolism 및 energy homeostasis를 조절한다.
4. PTP1B inhibitor 97개의 molecule을 이용해 QSAR 모델을 만들었으며, 기존의 Reference paper에서 제시한 QSAR 모델과 비교한 결과, **pyQSAR module을 이용한 QSAR 모델이 Reference보다 상대적으로 더 높게 나왔으며** R^2 과 Q^2 의 차이가 더 적게 나왔다.
5. 해당 QSAR model을 통해서 Diabetes과 Obesity에 대한 **신약을 개발하는 데에 도움을 줄 수 있다.**

Thank you