Electronic Supplementary Material (ESI) for RSC Advances.

Supplementary Information

Property-Oriented Design Framework for Rapid Discovery of Energetic Molecule Based on Small-Scale Labeled Dataset

Yunhao Xie [a], Yijing Liu [b], Renling Hu[a], Xu Lin[a], Jing Hu[a], Xuemei Pu [a,*]

[a] College of Chemistry, Sichuan University, Chengdu, 610064

[b] College of Computer Science, Sichuan University, Chengdu, 610064

* Corresponding Author

Xuemei Pu (xmpuscu@scu.edu.cn)

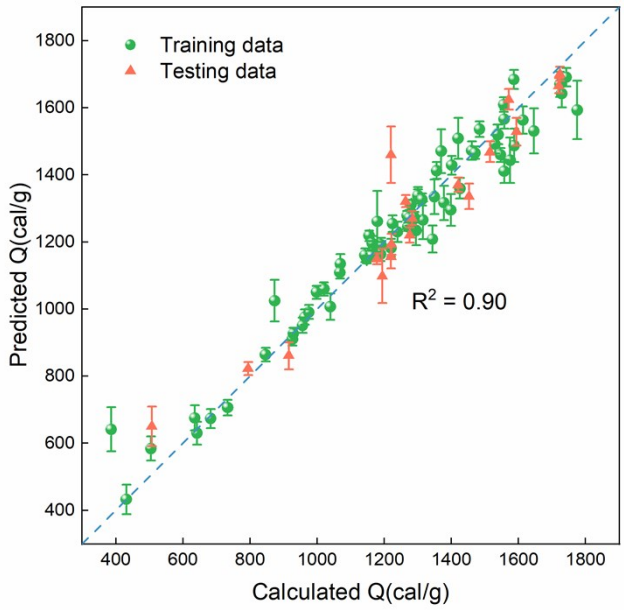

Fig. S1 The relationship between the average values predicted by KRR coupled with E-state+SOB+CDS over 20 repeated tests and the calculated values derived from the QM method and empirical equation for the heat of explosion. The standard deviation over the 20 repeated tests is marked by the error bar.

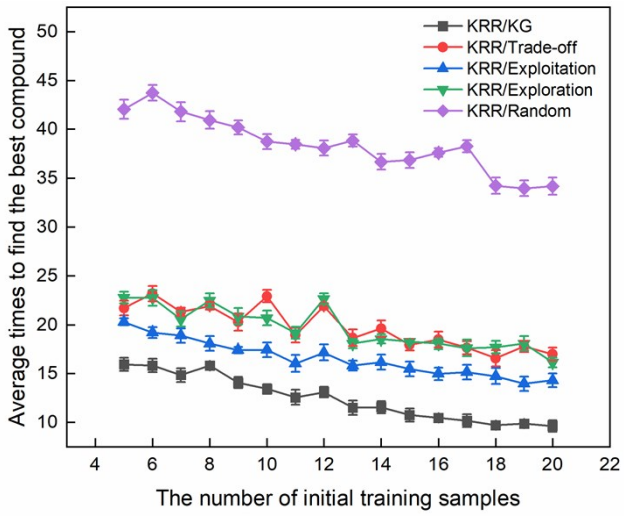

Fig. S2 Comparison of performance for the 10 regressor/optimizer combination pairs on different initial dataset randomly selected from the 88 samples labeled for the heat of explosion, derived from 20 repeated tests. The x axis represents the number of initial samples randomly selected from the 88 compounds to establish a statistical inference model. The y axis represents the average number of iterations required to find the best compound (i.e. the compound with the highest heat of explosion) in the 88 samples. The standard deviation of the results over the 20 repeated tests is marked by the error bar. The best regressor/optimizer combination pair is able to find the best compound at the least number of iterations. The results show that the combination of KRR/KG (continuous purple line) is the best regressor/optimizer combination pair.

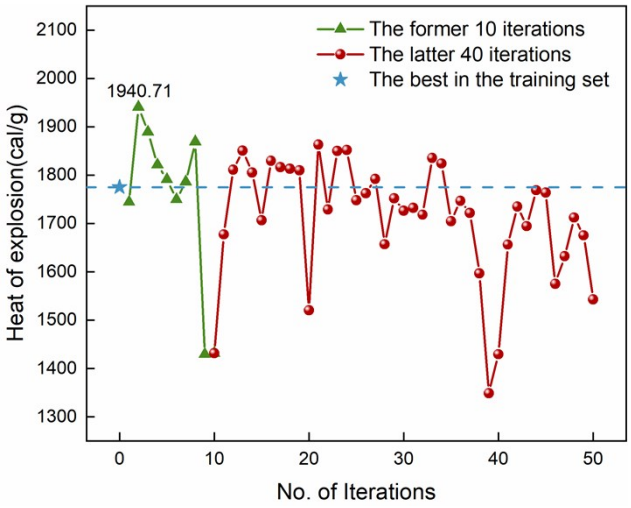

Fig. S3 The calculated heat of explosion derived from the QM calculation and empirical equation of 50 compounds selected from the first 50 iterations by means of the KRR/KG combination pair. The former 10 iterations and the latter 40 iterations are represented by green triangles and red spheres, respectively. Iteration 0 means that the training set is consisted of the initial 88 compounds, where the maximum value of the heat of explosion is 1775.19 cal/g.