Discovering New DNA Gyrase Inhibitors Using Machine Learning Approaches

Long Li¹, Xiu Le¹, Ling Wang^{2,3}, Qiong Gu¹, Huihao Zhou¹ and Jun Xu^{*1} ¹Research Center for Drug Discovery, School of Pharmaceutical Sciences, Sun Yat-Sen University, Guangzhou 510006, China

²Pre-Incubator for Innovative Drugs & Medicine, School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, China ³Guangdong Provincial Key Laboratory of Fermentation and Enzyme Engineering, School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, China

Content

Figure S1.5 μ M was the activity threshold.

Figure S2. The proportion selection of Training set and Test set.

Figure S3. Compare the performances of combinational NB models and fingerprintbased NB models for the tested compounds.

Figure S4. The Matthews correlation coefficient (C) versus the tree depth for the training set and test set.

Figure S5. The SCA-plot of the 488 compounds (red dots) and the known DNA gyrase inhibitors (blue dots).

Figure S6. SDS-PAGE results of (A) E.coil DNA GyraseA and (B) E.coil DNA GyraseB.

Figure S7. The activity distribution in the training and test sets.

Table S1. Detail information of the 137 compounds.

Table S2. Detail information of training set and test set.

Table S3. Detail information of 113 external tested compounds.

Table S4.Detail information of RCDD Database.

 Table S5.Detail information of 29 NB models and 16 RP models generated from descriptors or fingerprints.

Table S6.Detail information of 54 NB models and 324 RP models generated from the

 combinations of descriptors and fingerprints.

Table S7. Detail information of the compounds screened from the 11 best models.



Figure S1.MCC- and fingerprints-based cutoff value analysis.



Figure S2. The proportion selection of Training set and Test set.



Figure S3. Compare the performances of combinational NB models and

fingerprint-based NB models for the tested compounds.

MCC: Matthews correlation coefficient





Figure S4. The Matthews correlation coefficient (*C*) versus the tree depth for the training set and test set.



MCC: Matthews correlation coefficient

Figure S5. The SCA-plot of the 488 compounds (red dots) and the known DNA

gyrase inhibitors (blue dots).



Figure S6.SDS-PAGE results of (A) *E.coil* DNA GyraseA and (B) *E.coil* DNA GyraseB.



Figure S7. The activity distribution in the training set (blue) and Test set(red).