Supporting Information

SyntaLinker: Automatic Fragment Linking with Deep Conditional Transformer Neural Networks

Yuyao Yang,‡[ab] Shuangjia Zheng,‡[a] Shimin Su,[ab] Chao Zhao[a], Jun Xu,*[a] and Hongming Chen*[b]

E-mail: chen_hongming@grmh-gdl.cn; junxu@biochemomes.com

[a.] *Research Center for Drug Discovery, School of Pharmaceutical Sciences, Sun Yat-Sen University, 132 East Circle at University City, Guangzhou 510006, China.*
[b.] *Center of Chemistry and Chemical Biology, Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou 510530, China.*
‡ *Authors contributed equally.*

Table S1 The range of some key hyperparameters in SyntaLinker and the chosen ones for the best model are in bold

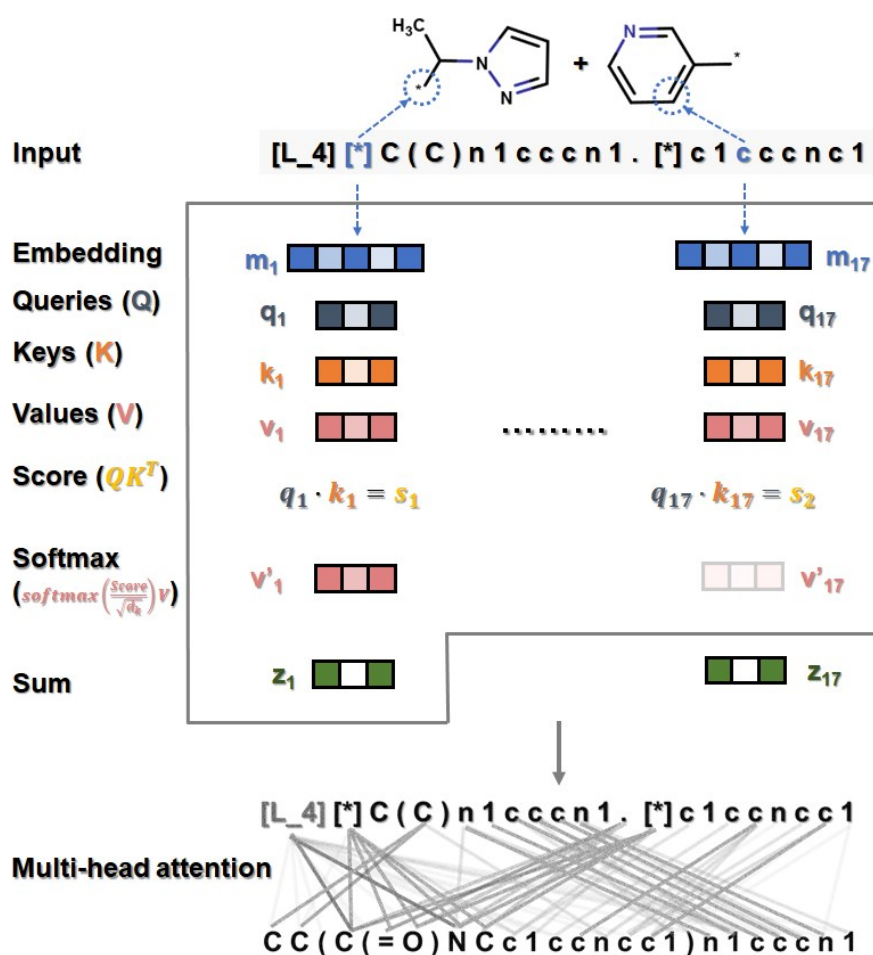| Hyperparameters	Search range and best value
Batch size	256, 502, 1024, 2048, **4096**
Hidden states	128, **256**, 512
Attention heads	2, 4, 6, **8**, 10
Optimizer	SGD, **Adam**, Adagrad
Starting learning rate	0.001 ~ 2 (**1**)
Drop out	0 ~ 1 (**0.1**)
Decay factor	0 ~ 1 (**0.5**)

Fig. S1 The graphical illustration of attention mechanism in SyntaLinker. The lines in the Multi-head attention map refers to some important attention weight between the output tokens and input tokens.