Supporting Information for

Autonomous Discovery of Unknown Reaction Pathways from Data by Chemical Reaction Neural Network

Weiqi Ji, Sili Deng *

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Corresponding Author: * E-mail: silideng@mit.edu (Sili Deng)

ORCID: Weiqi Ji: 0000-0002-7097-0219, Sili Deng: 0000-0002-3421-7414.

1. Case I: An Elementary Reaction Network Without Temperature Dependence

1.1 Governing Equations

The governing equations for Case I are shown in Eq. **S1.**

$$\frac{d[A]}{dt} = -2k_1[A]^2 - k_2[A]$$

$$\frac{d[B]}{dt} = k_1[A]^2 - k_4[B][D]$$

$$\frac{d[C]}{dt} = k_2[A] - k_3[C] \qquad \textbf{(S1)}$$

$$\frac{d[D]}{dt} = k_3[C] - k_4[B][D]$$

$$\frac{d[E]}{dt} = k_4[B][D]$$

2. Case II: Bio-diesel Production with Temperature Dependence

2.1 Governing Equations

The governing equations for Case II are shown in Eq. **S2.**

$$\frac{d[TG]}{dt} = -r_1$$

$$\frac{d[ROH]}{dt} = -r_1 - r_2 - r_3$$

$$\frac{d[DG]}{dt} = r_1 - r_2$$

$$\frac{d[MG]}{dt} = r_2 - r_3$$

$$\frac{d[GL]}{dt} = r_3$$ **(S2)**

$$\frac{d[R'CO_2R]}{dt} = r_1 + r_2 + r_3$$

$$r_1 = k_1[TG][ROH]$$

$$r_1 = k_2[DG][ROH]$$

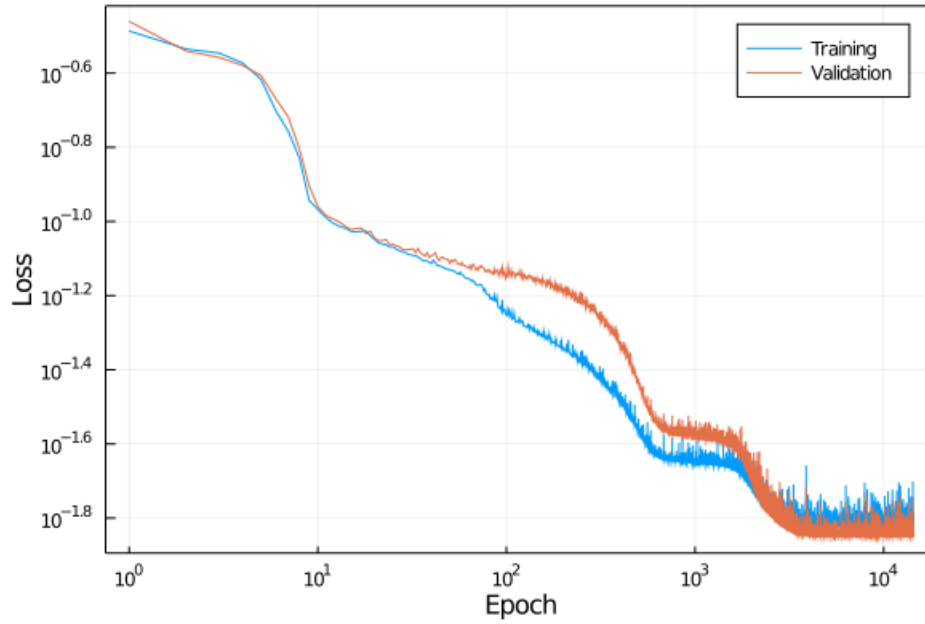$$r_1 = k_3[MG][ROH]$$

2.2 Loss Curves

Figure. S1. The typical evolution of loss functions with the number of epochs for Case II. Results shown correspond to the CRNN with three hidden nodes.
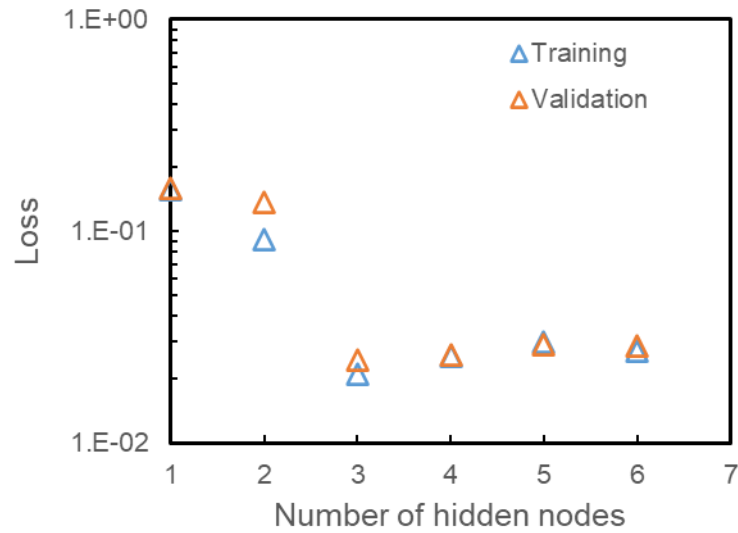
2.3 Incomplete Dataset

Figure. S2. The dependence of minimum loss functions for training and validation dataset with the number of hidden nodes.

3. Case III: An Enzyme Reaction Network

3.1 Governing Equations

The governing equations for Case III are shown in Eq. **S3.**

$$\frac{d[S]}{dt} = 0$$

$$\frac{d[MAP3K]}{dt} = -r_1 + r_5$$

$$\frac{d[MAP3K^*]}{dt} = r_1 - r_5$$

$$\frac{d[MAP2K]}{dt} = -r_2 + r_6$$

$$\frac{d[MAP2K^*]}{dt} = r_2 - r_6$$

$$\frac{d[MAPK]}{dt} = -r_3 + r_7$$

$$\frac{d[MAPK^*]}{dt} = r_3 - r_7$$

$$\frac{d[TF]}{dt} = -r_4 + r_8 \tag{S3}$$

$$\frac{d[TF^*]}{dt} = r_4 - r_8$$

$$r_1 = k_1[S][MAP3K]$$

$$r_2 = k_2[MAP3K^*][MAP2K]$$

$$r_3 = k_3[MAP2K^*][MAPK]$$

$$r_4 = k_4[MAPK^*][TF]$$

$$r_5 = k_5[MAP3K^*]$$

$$r_6 = k_6[MAP2K^*]$$

$$r_7 = k_7[MAPK^*]$$
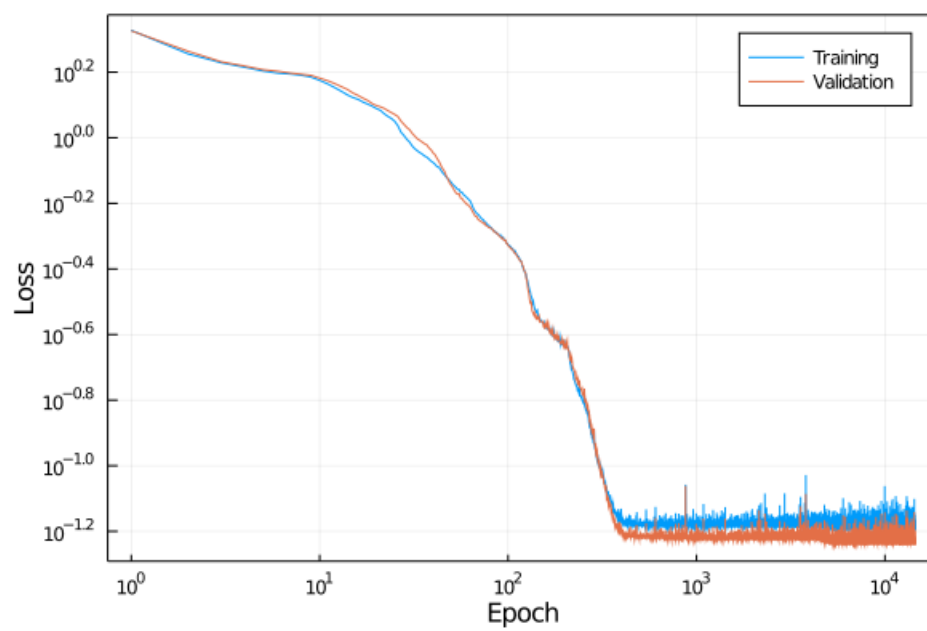
$$r_8 = k_8[TF^*]$$

3.2 Loss Curves

Figure S3. The typical evolution of loss functions with the number of epochs for Case III. Results shown correspond to the CRNN with eight hidden nodes.