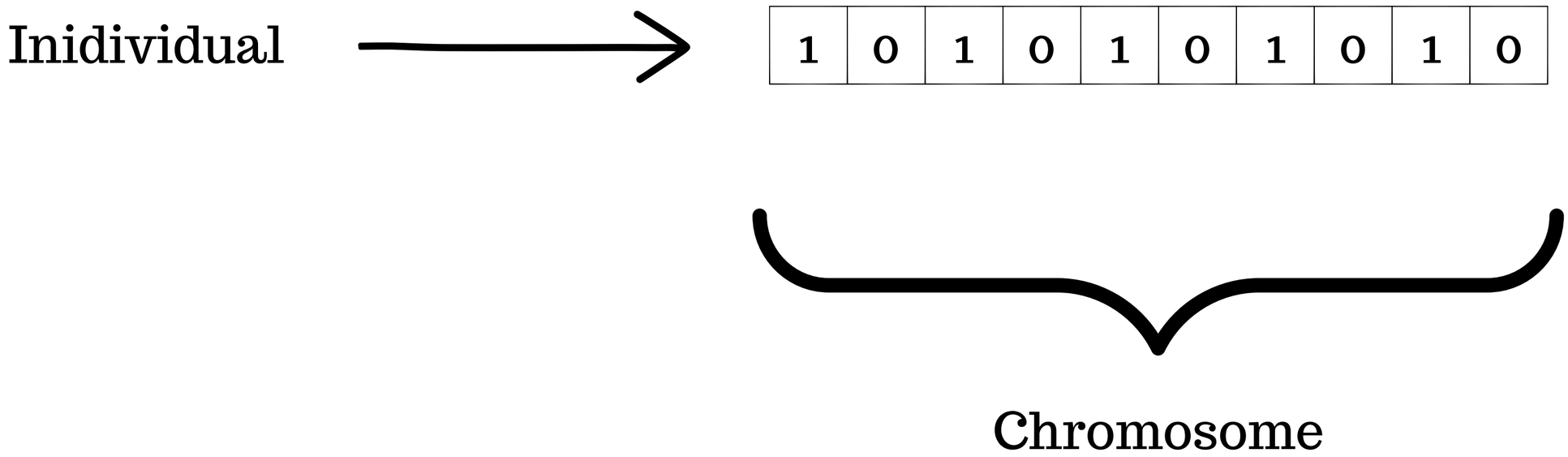
Generative Molecular Design

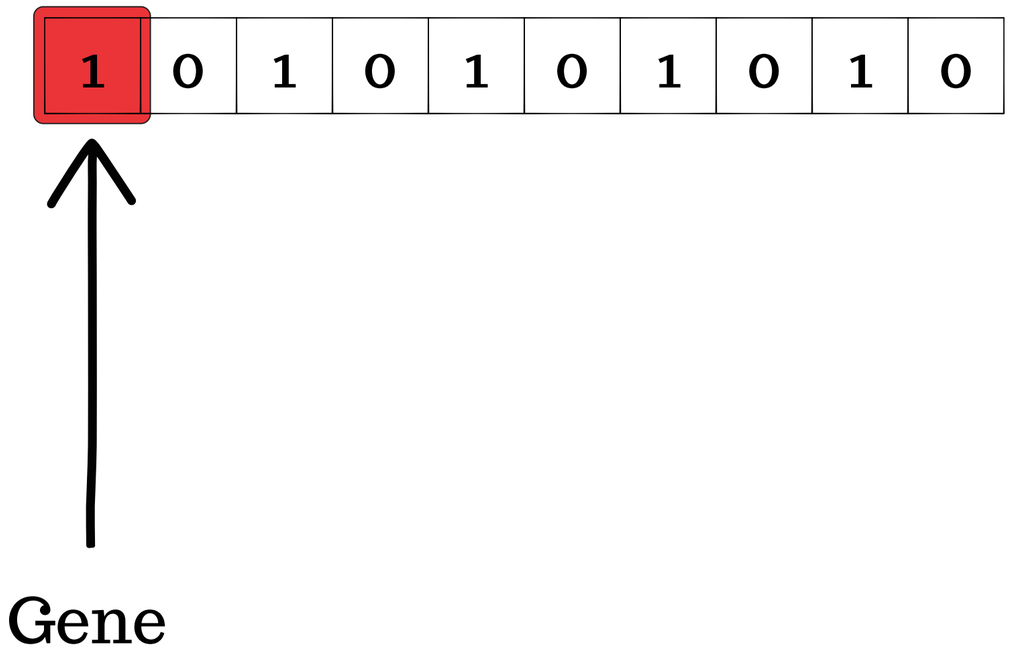
Canonical Genetic Algorithm

- Another form of random search
- Based on basic concept of evolution
  - Survival of the fittest
  - Population based
  - Sexual reproduction

Canonical Genetic Algorithm

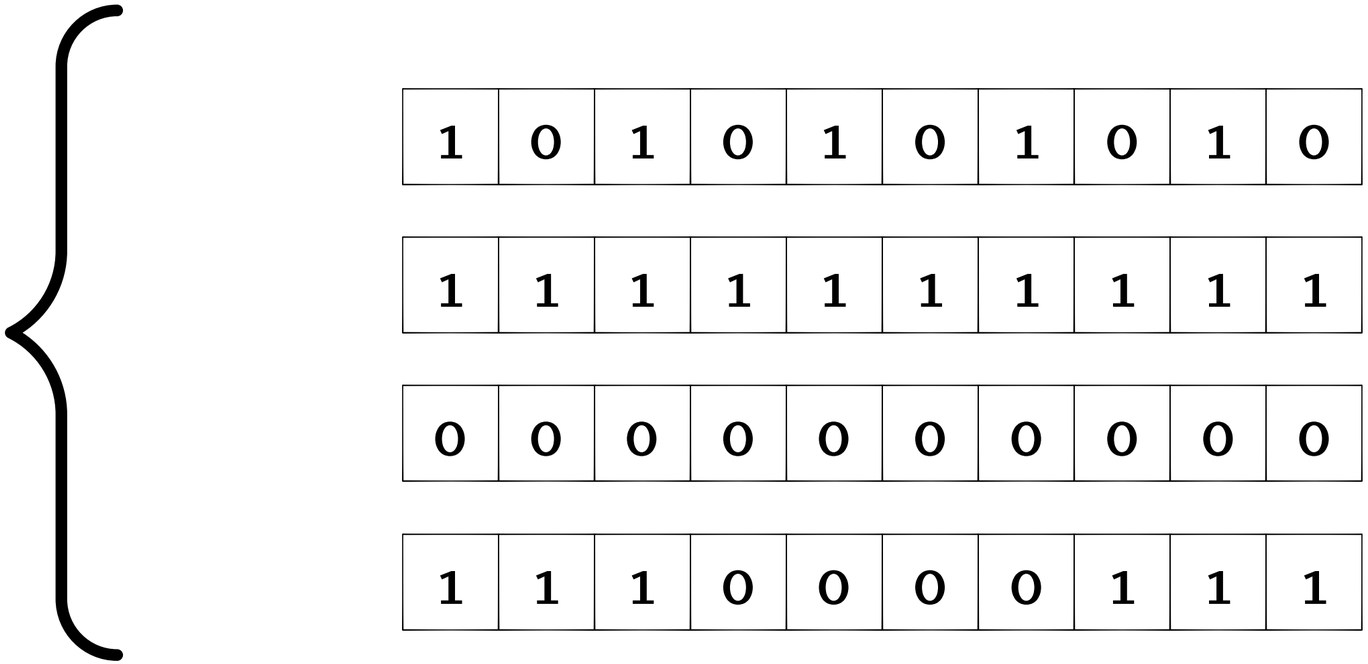


Canonical Genetic Algorithm

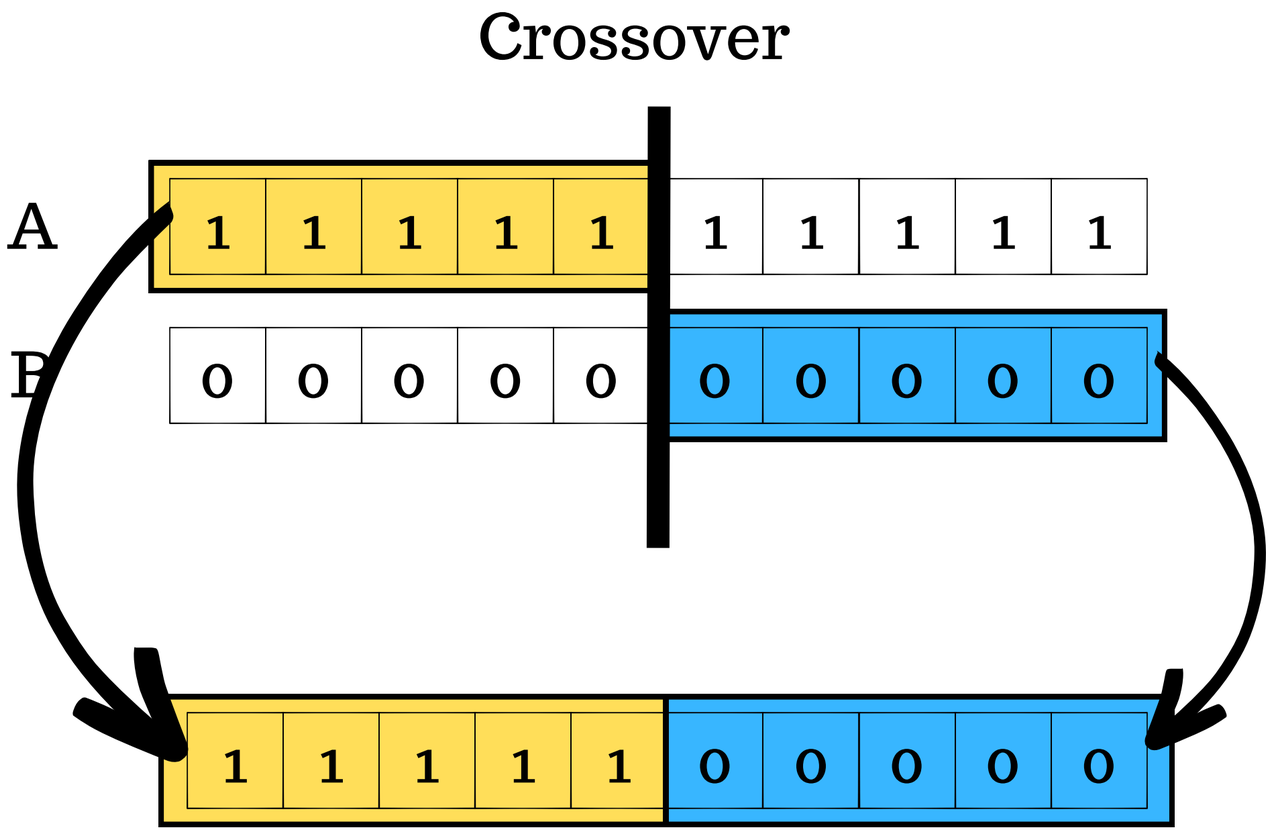


Canonical Genetic Algorithm

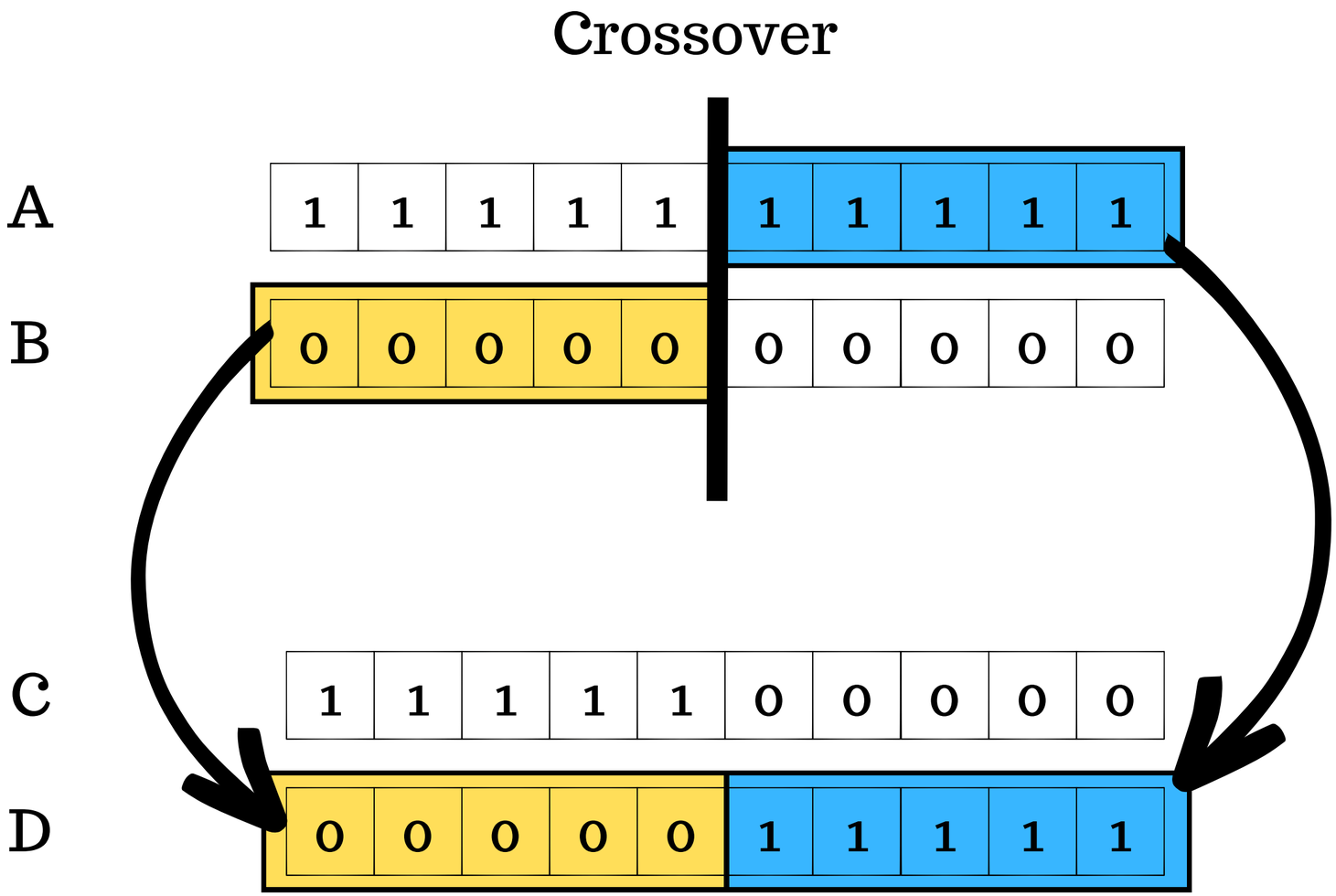
Population



Canonical Genetic Algorithm

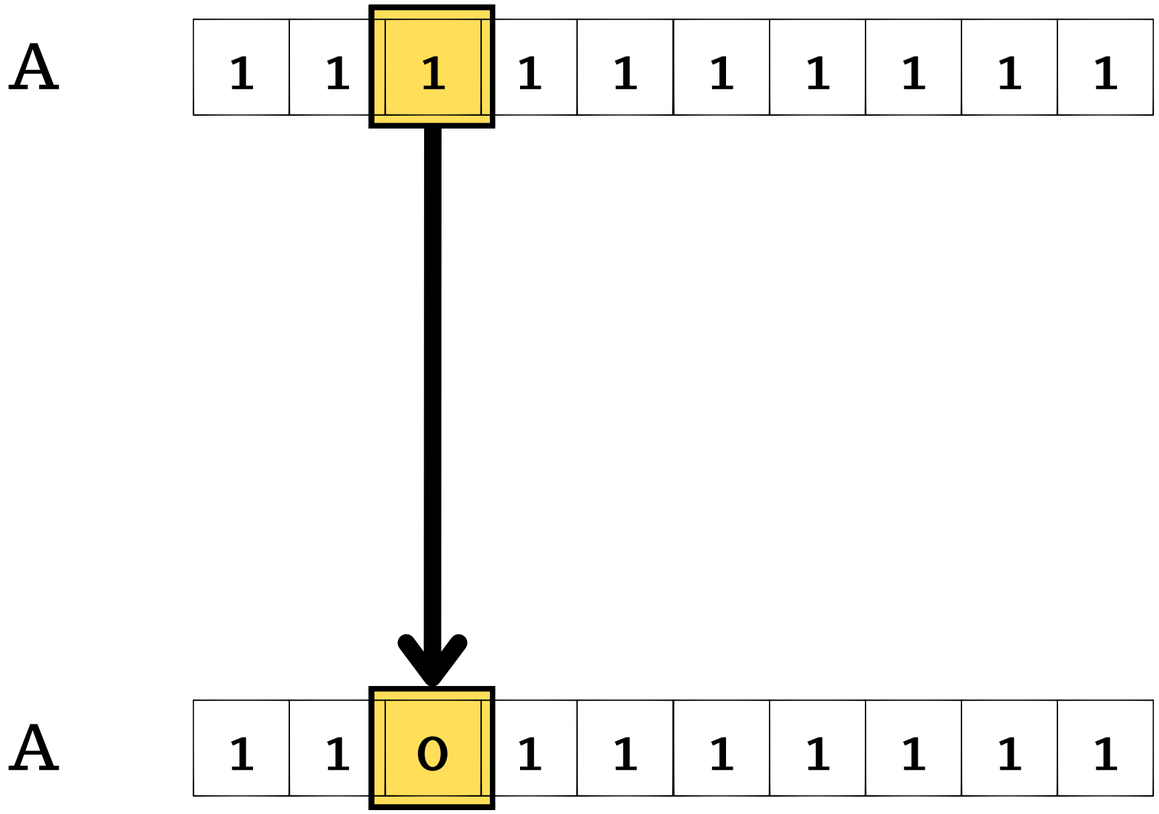


Canonical Genetic Algorithm



Canonical Genetic Algorithm

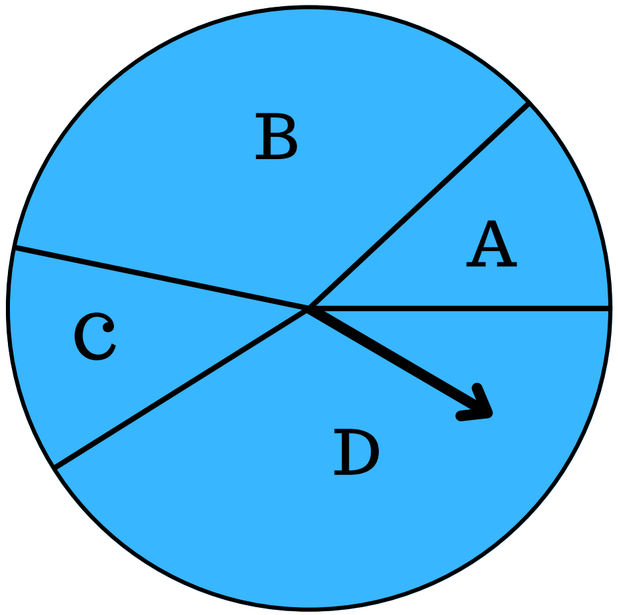
Mutation



Canonical Genetic Algorithm

Selection

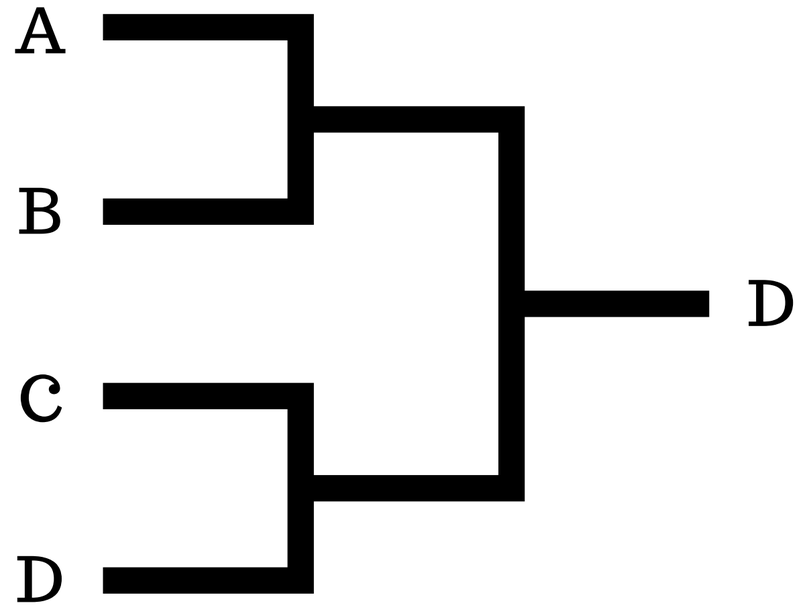
Roulette wheel



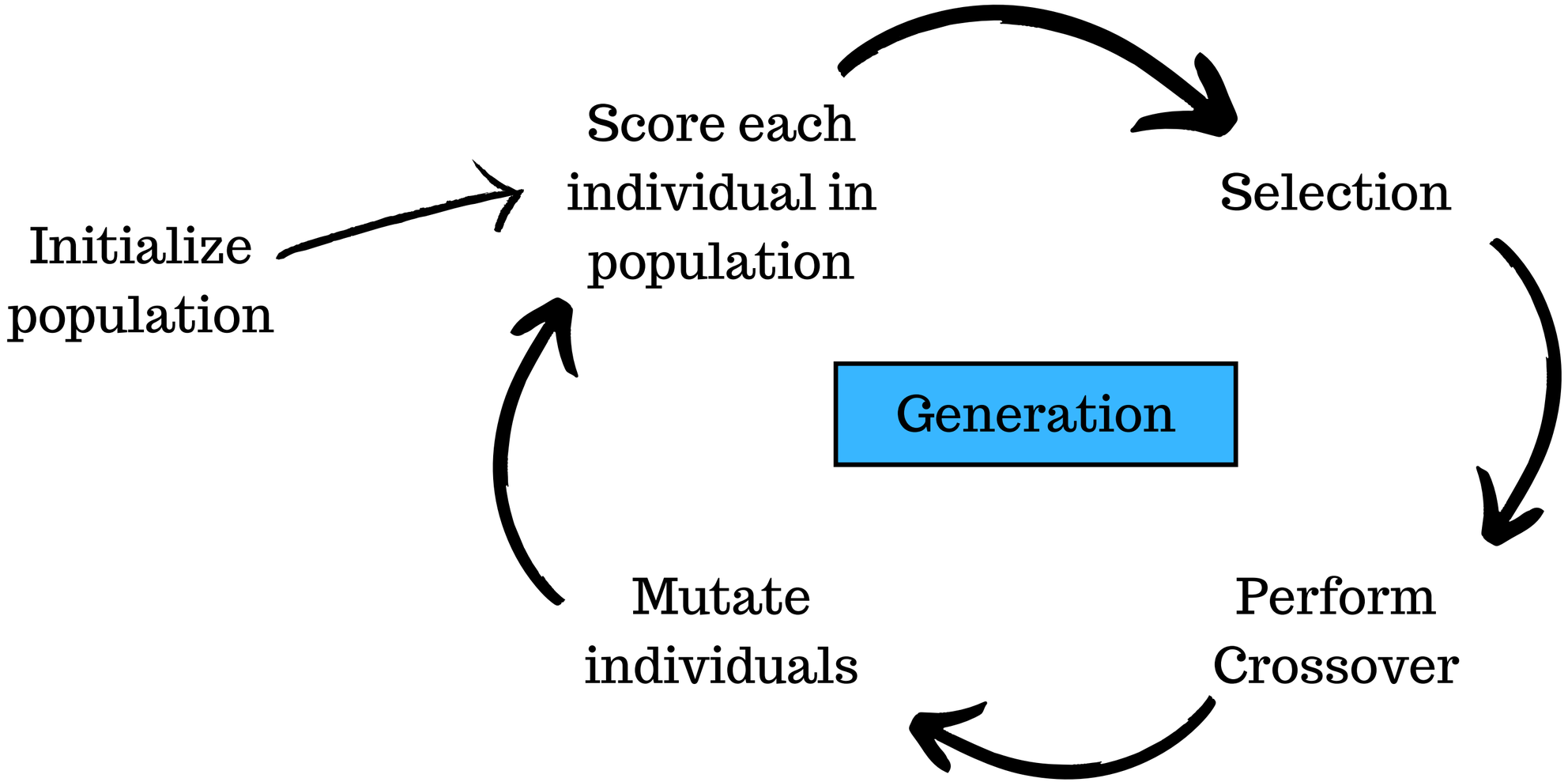
Rank Selection

ID	fitness	Rank
D	15	1
B	3	2
A	2.5	3

Tournament Selection



Canonical Genetic Algorithm

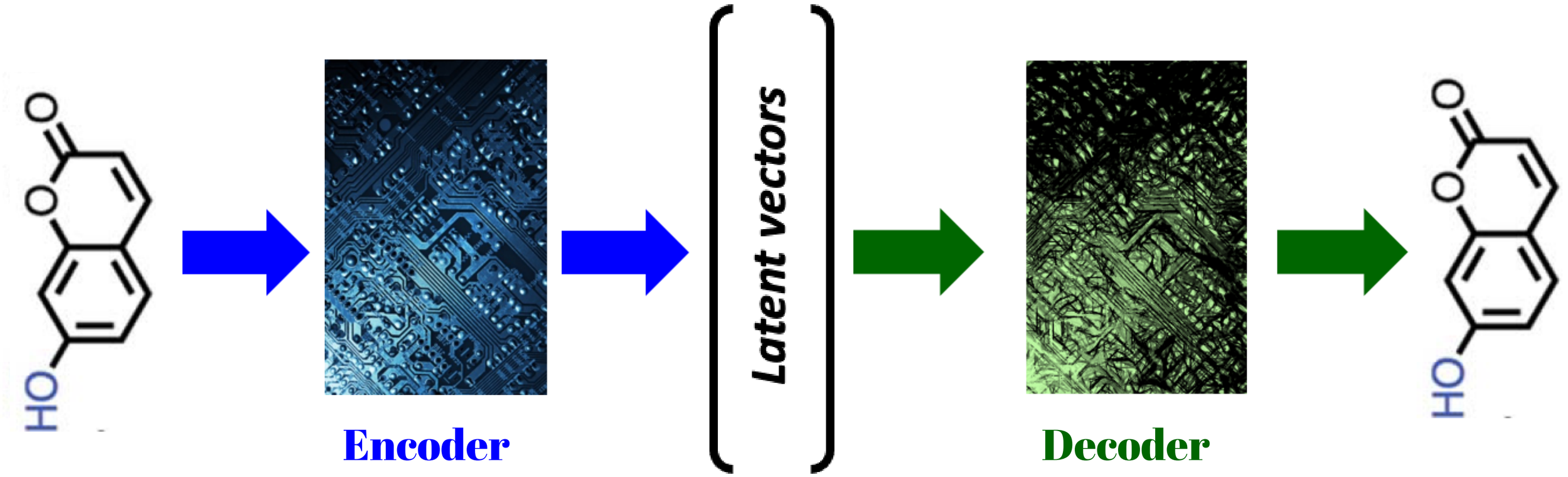


Generative Molecular Design

- Often follows the same general form as the typical GA
- One clear issue: Canonical GA requires a chromosomal representation of a molecule!
    - This is where latent space representations came in:

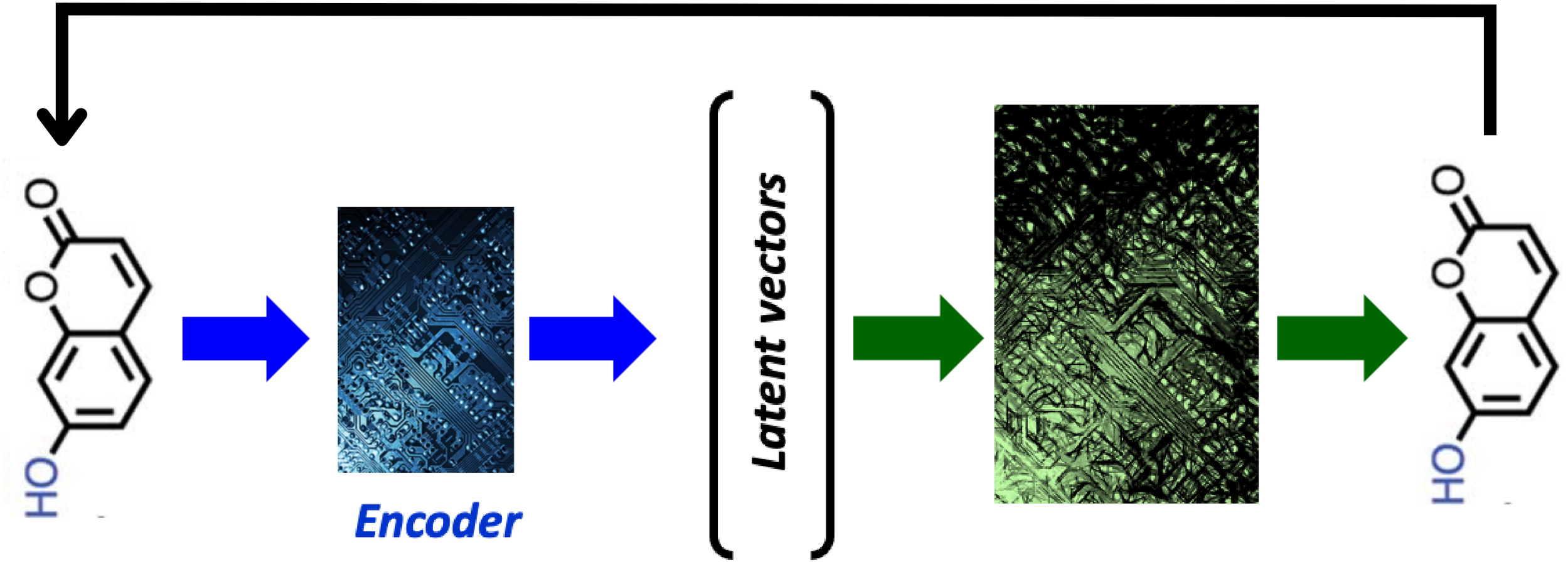
Generative Molecular Design

Autoencoder



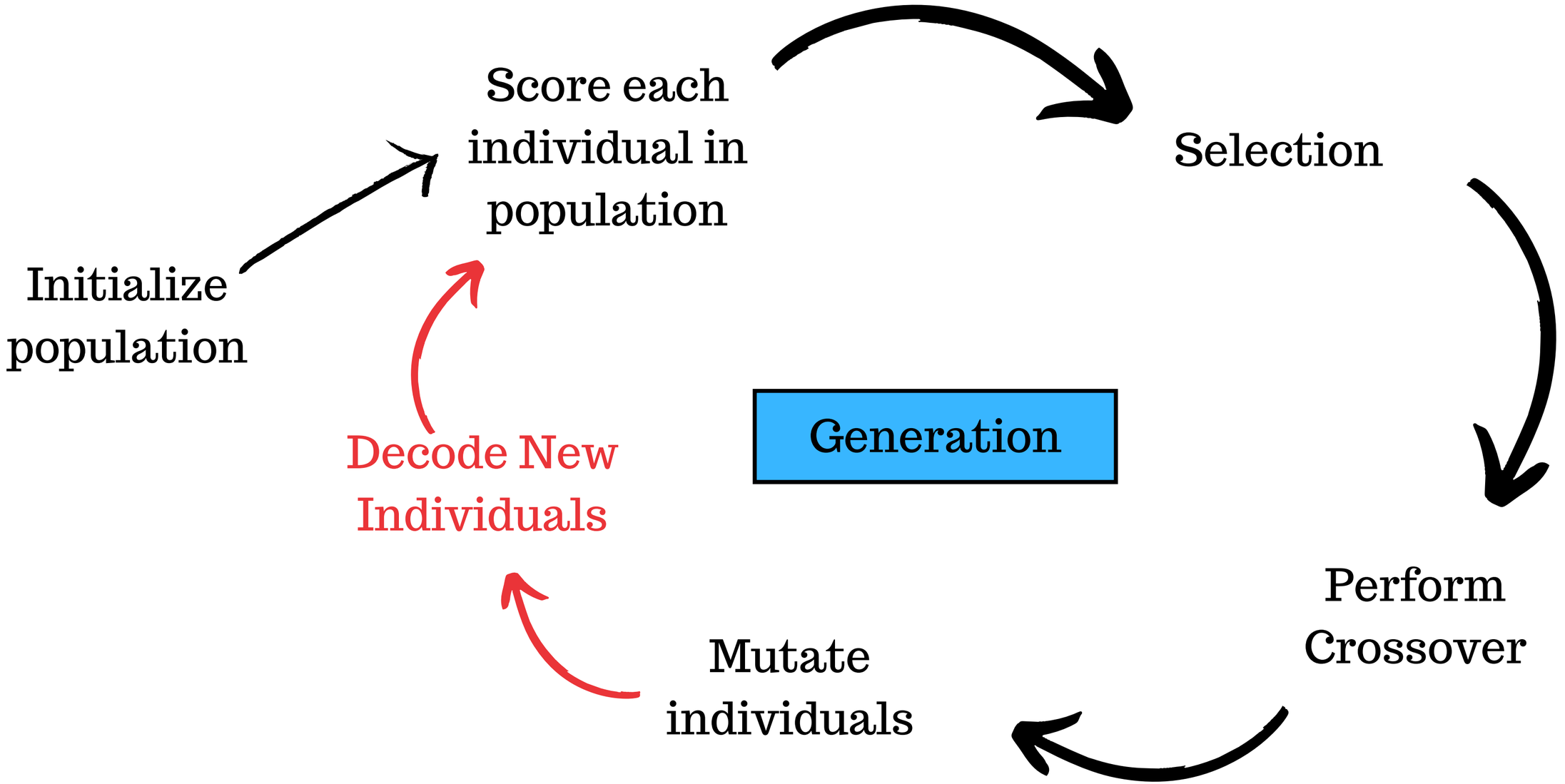
Generative Molecular Design

Hopefully the original and the decoded molecule match

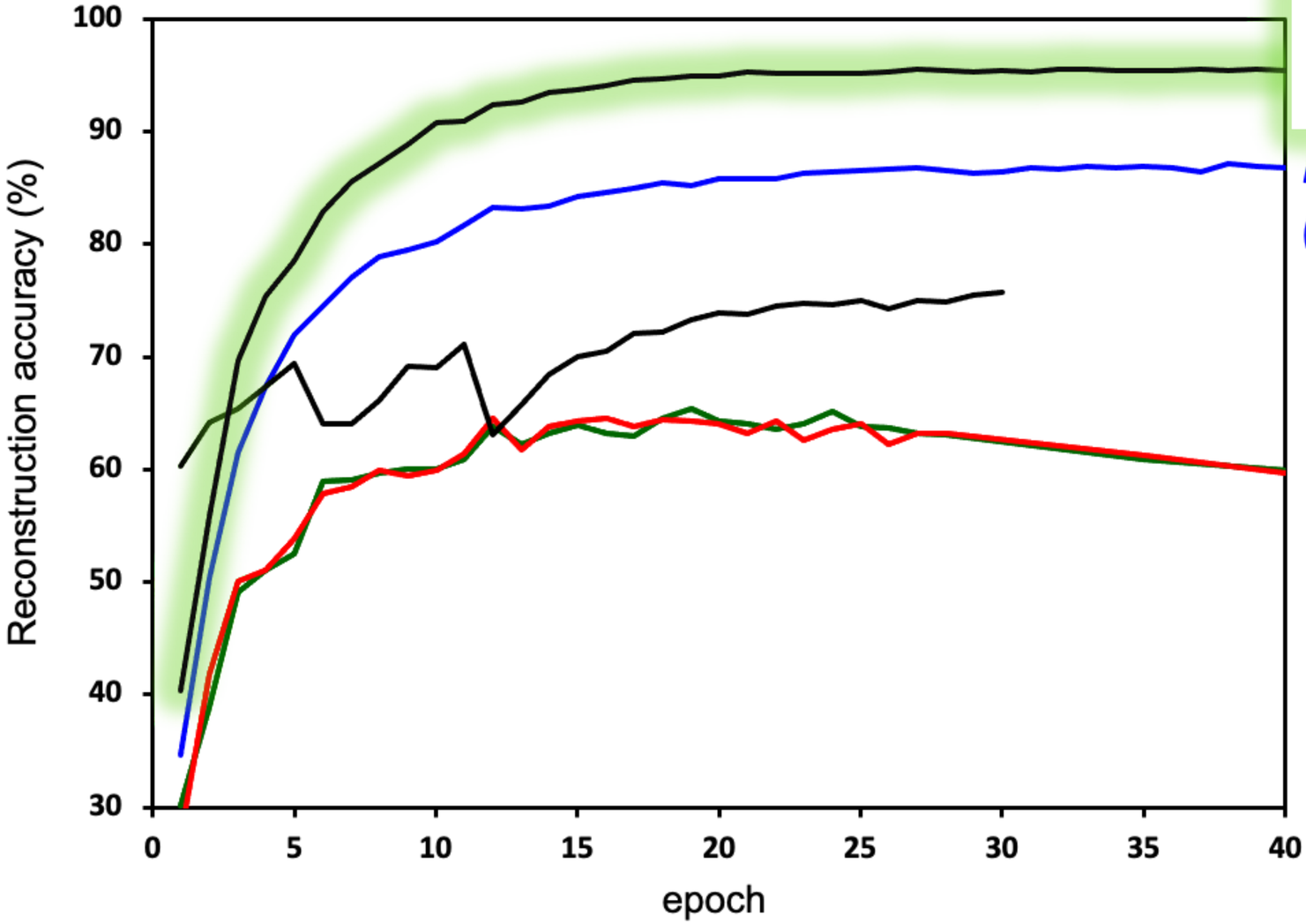


Autoencoder

Generative Molecular Design



Generative Molecular Design



Train on All: 96 %
(All ZINC 250K)

Fast-New: 87 %
(modified Model training)

Performance of code from original JTVAE paper: 77%

Fast-Default: 65 %
(suggested/default Model training process)

Generative Molecular Design

There are issues with the latent space:

- Chemical space is limited to the space "known" by the latent space
- Latent space does NOT understand chemistry!
- Reconstruction error will never be 100%
- The dataset the autoencoder is trained on effects the latent space and the kind of molecules that are likely to be created.