**Supplemental Information Guide**

Folders within the Directory:

**Generated\_Molecules\_With\_Similarity**:

- final\_similarity\_data.pdf: Data File of Generated Molecules originating from SRC Kinase Inhibitors and the known SRC Kinase Inhibitor it is closest to.

Columns: SRC Mol (Known SRC Kinase Inhibitor)

Similarity (similarity score)

Mol (Generated Novel Molecule)

smiles (SMILE string of generated molecule)

qed (QED value of generated molecule)

logP (logP value of generated molecule)

SAS (SAS value of generated molecule)

rings (number of rings for generated molecule)

kinase inhibition likeliness (kinase inhibition score of generated molecule)

- final\_similarity\_data\_kinase\_family (1).pdf: Data File of Similarity with Generated Molecules originating from various Kinase Inhibitor families and the known SRC Kinase Inhibitor it is closest to.

Columns: SRC Mol (Known SRC Kinase Inhibitor)

Similarity (similarity score of generated molecule to known SRC Kinase Inhibitor)

smiles (SMILE string of generated molecule)

Mol (Generated Novel Molecule)

Class (Originating Kinase Family):

0: ABL1

2: EGFR

3: CSF1R

4: FLT3

5: KDR

6: LCK

7: MAPK10

8: MAPK14

9: MET

qed (QED value of generated molecule)

logP (logP value of generated molecule)

SAS (SAS value of generated molecule)

rings (number of rings for generated molecule)

kinase inhibition likeliness (kinase inhibition score of generated molecule)

**Original\_Set\_of\_Generated\_Molecules:**

- Original\_Set.csv: Original set of generated molecules produced from kinase families

Columns: smiles (SMILE strings for generated molecules)

distance (how far the new encoding is from the original encoding)

count (number of each SMILE string produced)

mol (generated molecule image path)

Class (Originating Kinase Family):

0: ABL1

2: EGFR

3: CSF1R

4: FLT3

5: KDR

6: LCK

7: MAPK10

8: MAPK14

9: MET

qed (QED value of generated molecule)

logP (logP value of generated molecule)

SAS (SAS value of generated molecule)

- All\_Uniques.csv: unique set of generated molecules produced by originating families— created from Original\_Set,csv

Columns: same as above, but includes structural and chemical properties calculated by RDKit

smiles (SMILE strings for generated molecules)

distance (how far the new encoding is from the original encoding)

count (number of each SMILE string produced)

mol (generated molecule image path)

Class (Originating Kinase Family):

0: ABL1

2: EGFR

3: CSF1R

4: FLT3

5: KDR

6: LCK

7: MAPK10

8: MAPK14

9: MET

Characteristics:

rings

weight

csp3

hallkier count

laputeASA

alicarbo

alihetero

alirings

amide bonds

arocarbo

arohetero

arorings

stereocenter

bridgehead

hba

hbd

rotatable

- Over\_75\_Kinase.csv: generated molecules that have a kinase score of 0.75 or above

Columns: smiles (SMILE strings for generated molecules)

distance (how far the new encoding is from the original encoding)

count (number of each SMILE string produced)

mol (generated molecule image path)

Class (Originating Kinase Family):

0: ABL1

2: EGFR

3: CSF1R

4: FLT3

5: KDR

6: LCK

7: MAPK10

8: MAPK14

9: MET

Characteristics:

rings

weight

csp3

hallkier count

laputeASA

alicarbo

alihetero

alirings

amide bonds

arocarbo

arohetero

arorings

stereocenter

bridgehead

hba

hbd

rotatable

Kinase Inhibition Likeliness

- Top\_100\_Mol.csv: top 100 generated molecules with highest similarity to known SRC kinase inhibitors

Columns: Known Mol (SRC Known Kinase Inhibitor image path)

similarity (similarity score of generated molecule to known SRC Kinase Inhibitor)

smiles (SMILE strings for generated molecules)

distance (how far the new encoding is from the original encoding)

count (number of each SMILE string produced)

mol (generated molecule image path)

Class (Originating Kinase Family):

0: ABL1

2: EGFR

3: CSF1R

4: FLT3

5: KDR

6: LCK

7: MAPK10

8: MAPK14

9: MET

Characteristics:

rings

weight

csp3

hallkier count

laputeASA

alicarbo

alihetero

alirings

amide bonds

arocarbo

arohetero

arorings

stereocenter

bridgehead

hba

hbd

rotatable

Kinase Inhibition Likeliness

**Prelim\_Data\_Generated\_Molecules\_from\_SRC\_Inhibitors: Initial data from the original experiment- generation of novel molecules from SRC Kinase Inhibitors**

- Top\_100\_Cluster\_1.xlsx: Top 100 Generated Molecules from cluster 1 for initial experiments- generation of novel molecules from SRC Kinase Inhibitors

Columns: Mol (generated molecule image)

smiles (SMILE strings for generated molecules)

distance (how far the new encoding is from the original encoding)

count (number of each SMILE string produced)

Characteristics:

rings

weight

csp3

hallkier count

laputeASA

alicarbo

alihetero

alirings

amide bonds

arocarbo

arohetero

arorings

stereocenter

bridgehead

hba

hbd

rotatable

Kinase Inhibition Likeliness

- Top\_100\_Cluster\_2.xlsx: Top 100 Generated Molecules from cluster 2 for initial experiments- generation of novel molecules from SRC Kinase Inhibitors

Columns: Mol (generated molecule image)

smiles (SMILE strings for generated molecules)

distance (how far the new encoding is from the original encoding)

count (number of each SMILE string produced)

Characteristics:

rings

weight

csp3

hallkier count

laputeASA

alicarbo

alihetero

alirings

amide bonds

arocarbo

arohetero

arorings

stereocenter

bridgehead

hba

hbd

rotatable

Kinase Inhibition Likeliness

- 2\_Cluster\_Top\_100\_1.csv & 2\_Cluster\_Top\_100\_2.csv: csv files of excel sheets

**Graphs\_OF\_Chemical\_Proprties:** graphs of chemical properties — average values of properties for generated molecules by originating families

- alicarbo\_bar: average aliphatic carbocycles of generated molecules by originating family

- alihetero\_bar: average aliphatic heterocycles of generated molecules by originating family

- alirings\_bar: average aliphatic rings of generated molecules by originating family

- arocarbo\_bar: average aromatic carbocycles of generated molecules by originating family

- arohetero\_bar: average aromatic heterocycles of generated molecules by originating family

- arorings\_bar: average aromatic rings of generated molecules by originating family

- csp3\_bar: average CSP3 value of generated molecules by originating family

- hallkier\_bar: average hallkier count of generated molecules by originating family

- labuteasa\_bar: average labuteasa value of generated molecules by originating family

- weight\_bar: average molecular weight of generated molecules by originating family

- rotatable\_bar: average rotatable bonds of generated molecules by originating family

**Histogram\_Distribution\_Of\_Chemical\_Properties**: Histogram distribution of properties for all molecules in the kinase families used for the training set- initial chemical analysis