What is Entrez in Biopython?

Entrez is a programming interface provided by **NCBI** that allows you to:

- Search biological databases
- Download sequences
- Fetch data using Accession IDs

In simple words:
Entrez = Bridge between Python and NCBI databases

Why Entrez is Important for Bioanalysts?

Using Entrez, you can:

- Automatically download sequences
- Avoid manual copy-paste
- Work with large datasets
- Perform real bioinformatics analysis

Step 1: Import Entrez Module

```
from Bio import Entrez
```

Step 2: Set Email (VERY IMPORTANT)

NCBI requires your email for tracking usage.

```
Entrez.email = "your_email@example.com"
```

Without this, NCBI may block your requests.

Step 3: Search a Database (Entrez.esearch)

Example: Search gene sequences in NCBI Nucleotide database

```
handle = Entrez.esearch(
    db="nucleotide",
    term="BRCA1[Gene] AND Homo sapiens[Organism]"
)
result = Entrez.read(handle)
print(result)
```

Output contains:

- Count of results
- List of IDs (GI / accession numbers)

Step 4: Fetch Sequence Using Accession ID (Entrez.efetch)

Example: Fetch DNA sequence in FASTA format

```
handle = Entrez.efetch(
    db="nucleotide",
    id="NM_007294",
    rettype="fasta",
    retmode="text"
)
sequence = handle.read()
print(sequence)
```

This downloads the sequence directly from NCBI.

Step 5: Save Downloaded Sequence to a File

```
with open("sequence.fasta", "w") as file:
    file.write(sequence)
```

Step 4: Fetch Sequence Using Accession ID (Entrez.efetch)

Example: Fetch DNA sequence in FASTA format

```
handle = Entrez.efetch(
    db="nucleotide",
    id="NM_007294",
    rettype="fasta",
    retmode="text"
)
sequence = handle.read()
print(sequence)
```

This downloads the sequence directly from NCBI.

Step 5: Save Downloaded Sequence to a File

```
with open("sequence.fasta", "w") as file:
    file.write(sequence)
```

Fetch GenBank File Using Accession ID

```
handle = Entrez.efetch(
    db="nucleotide",
    id="NM_007294",
    rettype="gb",
    retmode="text"
)
genbank_data = handle.read()
print(genbank_data)
```

This file contains:

- Sequence
- Features
- Annotations
- References