

- Sau khi tải về các file FASTA, mỗi file chứa một chuỗi protein từ NCBI, nhóm tiến hành hợp nhất tất cả các file này thành một file FASTA lớn để tiện cho việc chạy BLAST.

```
cat *.fasta > all_proteins.fasta
```

- Để có thể thực hiện so khớp bằng *blastp*, cần chuyển đổi tập hợp các chuỗi protein trong *all_proteins.fasta* thành một cơ sở dữ liệu BLAST có cấu trúc tối ưu.

```
makeblastdb -in all_proteins.fasta -dbtype prot -out protein_db
```

- Chuẩn bị file dữ liệu FASTA cần so sánh. Một chuỗi protein mới được tạo thành file *query.fasta*. File này có cấu trúc chuẩn định dạng FASTA, có thể đến từ cơ sở dữ liệu mới, hoặc là chuỗi được người dùng cung cấp cần so khớp với cơ sở dữ liệu đã có để tìm protein tương đồng.
- Thực hiện lệnh BLAST

```
blastp -query query.fasta -db protein_db -out result.txt -outfmt 6
```

Sau khi lệnh thực hiện xong, kết quả được lưu tại file *result.txt*. Có thể xem kết quả trực tiếp bằng lệnh

```
cat result.txt
```