

```
python3 protein_search.py \
-d protein_vectors.fvecs \
-q targets.fasta \
-s swissprot_50k.fasta \
-o benchmarks/res_ivf_tuned.txt \
-method ivf \
-N 50 \
--ivf_k 200 \
--ivf_probe 20
```

Query Protein: A0A009I3Y5

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
IVF-Flat	11.479	0	0.00
BLAST (Ref)	1.285	0	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: IVF-Flat

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp P44033 Y665_HAEIN	0.36	1%	No	Possible false positive
2	sp A3DA85 SELU_SHEB5	0.41	3%	No	Possible false positive
3	sp Q4QLU9 RLMC_HAEI8	0.41	0%	No	Possible false positive
4	sp Q8EJX6 SELO_SHEON	0.41	1%	No	Possible false positive

5	sp P43954 Y152_HAEIN	0.44	1%	No	Possible false positive
6	sp A3D8E4 RSMC_SHEB5	0.44	1%	No	Possible false positive
7	sp Q04N32 RECO_STRP2	0.44	1%	No	Possible false positive
8	sp Q74Q56 SGRR_YERPE	0.44	1%	No	Possible false positive
9	sp Q1CGK9 SULA_YERPN	0.44	22%	No	Possible false positive
10	sp Q8X8V8 MTFA_ECO57	0.44	1%	No	Possible false positive

Query Protein: A0A009PCK4

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
IVF-Flat	11.479	0	0.00
BLAST (Ref)	0.309	3	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: IVF-Flat

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp P44033 Y665_HAEIN	0.36	1%	No	Possible false positive
2	sp A3DA85 SELU_SHEB5	0.41	3%	No	Possible false positive
3	sp Q4QLU9 RLMC_HAEI8	0.41	0%	No	Possible false positive
4	sp Q8EJX6 SELO_SHEON	0.41	1%	No	Possible false positive
5	sp P43954 Y152_HAEIN	0.44	1%	No	Possible false positive

6	sp Q04N32 RECO_STRP2	0.44	1%	No	Possible false positive
7	sp A3D8E4 RSMC_SHEB5	0.44	1%	No	Possible false positive
8	sp Q74Q56 SGRR_YERPE	0.44	1%	No	Possible false positive
9	sp A2BRQ6 SASA_PROMS	0.44	1%	No	Possible false positive
10	sp Q8X8V8 MTFA_ECO57	0.44	1%	No	Possible false positive

Query Protein: A0A001

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
--------	---------------	-----	-------------------

IVF-Flat	11.479	0	0.06
----------	--------	---	------

BLAST (Ref)	0.604	1	1.00 (ref)
-------------	-------	---	------------

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: IVF-Flat

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
------	-------------	---------	----------------	-----------------	-------------

1	sp Q20Z38 NDVA_RHOPB	0.34	1%	Yes	-
2	sp Q51719 YCOBA_PROFR	0.40	1%	No	Possible false positive
3	sp Q8P8W4 MSBA_XANCP	0.40	1%	Yes	-
4	sp Q21BF6 MODC_RHOPB	0.42	1%	No	Possible false positive
5	sp P9WGD8 FTSY_MYCTO	0.45	1%	No	-
6	sp P9WMX0 MFTF_MYCTO	0.46	2%	No	-

7	sp Q5P6D5 MACB_AROAE	0.46	1%	No	-
8	sp P9WIL0 PANE_MYCTO	0.46	1%	No	-
9	sp A1UG51 POTA_MYCSK	0.47	1%	No	-
10	sp Q67JX3 ECFA1_SYMTH	0.47	1%	No	-

Query Protein: A0A002

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
--------	---------------	-----	-------------------

IVF-Flat	11.479	0	0.10
----------	--------	---	------

BLAST (Ref)	0.555	1	1.00 (ref)
-------------	-------	---	------------

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: IVF-Flat

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
------	-------------	---------	----------------	-----------------	-------------

1	sp Q20Z38 NDVA_RHOPB	0.36	2%	Yes	-
2	sp Q8P8W4 MSBA_XANCP	0.41	1%	Yes	-
3	sp Q51719 YCOBA_PROFR	0.42	1%	No	Possible false positive
4	sp Q21BF6 MODC_RHOPB	0.45	1%	No	-
5	sp Q67JX3 ECFA1_SYMTH	0.47	1%	No	-
6	sp Q5P6D5 MACB_AROAE	0.48	1%	No	-
7	sp Q4FS42 MSBA_PSYA2	0.49	1%	Yes	-

8	sp Q3KJQ7 TAUB_PSEPF	0.50	1%	No	-
9	sp C5C0T0 MENG_BEUC1	0.50	2%	No	-
10	sp A1UG51 POTA_MYCSK	0.50	1%	No	-

Query Protein: A0A009HL96

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
IVF-Flat	11.479	0	0.08
BLAST (Ref)	0.257	3	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: IVF-Flat

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp Q5HI09 GRAR_STAAC	0.18	1%	Yes	-
2	sp G3XCY6 GLTR_PSEAE	0.42	3%	Yes	-
3	sp A6QJN2 NREB_STAAE	0.48	1%	No	-
4	sp P0ACS7 RPIR_ECOLI	0.51	1%	No	-
5	sp O07586 CUER_BACSU	0.51	14%	No	-
6	sp O34994 CCPN_BACSU	0.52	1%	No	-
7	sp A5VUU3 DIVK_BRUO2	0.52	25%	Yes	Twilight zone
8	sp B7LLW0 NSRR_ESCF3	0.52	15%	No	-

9	sp Q2FY80 SRRB_STAA8	0.52	1%	No	-
10	sp Q0AYL3 CHEB1_SYNWW	0.53	1%	Yes	-

Query Protein: A0A009HLV9

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
--------	---------------	-----	-------------------

IVF-Flat	11.479	0	0.08
----------	--------	---	------

BLAST (Ref)	0.413	2	1.00 (ref)
-------------	-------	---	------------

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: IVF-Flat

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
------	-------------	---------	----------------	-----------------	-------------

1	sp Q3K999 SFNR_PSEPF	0.23	2%	Yes	-
2	sp A8A3I6 NORR_ECOHS	0.24	1%	Yes	-
3	sp B5F3J7 RUVB_SALA4	0.25	0%	No	Possible false positive
4	sp P0A812 RUVB_ECOLI	0.26	0%	No	Possible false positive
5	sp P71229 HYFR_ECOLI	0.26	1%	Yes	-
6	sp Q15RN6 RUVB_PSEA6	0.26	0%	No	Possible false positive
7	sp Q7WGB3 RUVB_BORBR	0.27	0%	No	Possible false positive
8	sp Q9KU18 CLPB_VIBCH	0.28	1%	No	Possible false positive
9	sp P0DF50 RUVB_STRP3	0.28	1%	No	Possible false positive

10 | sp|A4STH5|NORR\_AERS4 | 0.28 | 1% | Yes | -

Query Protein: A0A009HN45

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

-----  
Method | Time/query(s) | QPS | Recall@N vs BLAST

-----  
IVF-Flat | 11.479 | 0 | 0.09

BLAST (Ref) | 0.611 | 1 | 1.00 (ref)  
-----

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: IVF-Flat

Rank | Neighbor ID | L2 Dist | BLAST Identity | In BLAST Top-N? | Bio comment

-----  
-----  
1 | sp|A8GSU6|SECA\_RICRS | 0.17 | 1% | No | Possible false positive  
2 | sp|O84707|SECA\_CHLTR | 0.19 | 1% | No | Possible false positive  
3 | sp|Q03ZQ8|SECA\_LEUMM | 0.20 | 1% | No | Possible false positive  
4 | sp|Q6MEX9|SECA\_PARUW | 0.20 | 2% | No | Possible false positive  
5 | sp|C0PWW7|UVRB\_SALPC | 0.21 | 0% | No | Possible false positive  
6 | sp|Q04J70|SECA\_STRP2 | 0.21 | 1% | No | Possible false positive  
7 | sp|Q15Q25|SECA\_PSEA6 | 0.21 | 2% | No | Possible false positive  
8 | sp|Q21MG1|SECA\_SACD2 | 0.21 | 1% | No | Possible false positive  
9 | sp|Q47EI2|UVRB\_DECAR | 0.21 | 0% | No | Possible false positive  
10 | sp|Q5HHR0|UVRB\_STAAC | 0.21 | 0% | No | Possible false positive

Query Protein: A0A009HQC9

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
--------	---------------	-----	-------------------

IVF-Flat	11.479	0	0.02
----------	--------	---	------

BLAST (Ref)	0.650	1	1.00 (ref)
-------------	-------	---	------------

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: IVF-Flat

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
------	-------------	---------	----------------	-----------------	-------------

1	sp Q48LP5 RAPA_PSE14	0.13	2%	Yes	-
2	sp C0PWW7 UVRB_SALPC	0.21	0%	No	Possible false positive
3	sp B7I8J5 SECA_ACIB5	0.21	1%	No	Possible false positive
4	sp B7NA78 UVRB_ECOLU	0.22	0%	No	Possible false positive
5	sp A1W465 SECA_ACISJ	0.22	1%	No	Possible false positive
6	sp Q21MG1 SECA_SACD2	0.22	1%	No	Possible false positive
7	sp Q15Q25 SECA_PSEA6	0.23	1%	No	Possible false positive
8	sp O84707 SECA_CHLTR	0.24	1%	No	Possible false positive
9	sp Q6MEX9 SECA_PARUW	0.24	2%	No	Possible false positive
10	sp Q03ZQ8 SECA_LEUMM	0.24	1%	No	Possible false positive



Query Protein: A0A009HPM0

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
IVF-Flat	11.479	0	0.02
BLAST (Ref)	0.492	2	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: IVF-Flat

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp A3DA85 SELU_SHEB5	0.31	1%	No	Possible false positive
2	sp A8GKD1 G6PI_SERP5	0.32	0%	No	Possible false positive
3	sp B1LNS5 TRMA_ECOSM	0.32	1%	No	Possible false positive
4	sp Q07YK6 PURT_SHEFN	0.32	1%	Yes	-
5	sp Q7MIG6 DXR_VIBVY	0.32	1%	No	Possible false positive
6	sp A3D8E4 RSMC_SHEB5	0.33	1%	No	Possible false positive
7	sp B2K9Z2 HCP_YERPB	0.33	1%	No	Possible false positive
8	sp B7I2C3 FMT_ACIB5	0.33	0%	No	Possible false positive
9	sp A1AJ76 MUTL_ECOK1	0.33	2%	No	Possible false positive
10	sp A3M3B1 SDHD_ACIBT	0.33	1%	No	Possible false positive

Query Protein: A0A009IB02

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
IVF-Flat	11.479	0	0.09
BLAST (Ref)	0.744	1	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: IVF-Flat

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp P10802 ODP2_AZOVI	0.28	1%	Yes	-
2	sp P29337 BCCP_STRMU	0.53	11%	Yes	-
3	sp Q0RRT2 RL7_FRAAA	0.60	13%	No	-
4	sp P45523 FKBA_ECOLI	0.61	1%	No	-
5	sp B5RCZ3 RODZ_SALG2	0.61	1%	No	-
6	sp Q73SE8 RL7_MYCPA	0.61	12%	No	-
7	sp Q7NQE5 RL7_CHRVO	0.61	15%	No	-
8	sp Q9UUZ6 RLA2_ASPFU	0.62	12%	No	-
9	sp B1JU13 RL7_BURO0	0.62	5%	No	-
10	sp Q60A07 RL7_METCA	0.62	14%	No	-

Query Protein: A0A010Q3W2

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
--------	---------------	-----	-------------------

IVF-Flat	11.479	0	0.05
----------	--------	---	------

BLAST (Ref)	0.492	2	1.00 (ref)
-------------	-------	---	------------

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: IVF-Flat

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
------	-------------	---------	----------------	-----------------	-------------

1	sp A2RAG5 ATG18_ASPNC	0.28	1%	No	Possible false positive
2	sp Q9S7I8 CDC22_ARATH	0.29	0%	No	Possible false positive
3	sp O82266 SWA1_ARATH	0.30	2%	Yes	-
4	sp Q8TEB1 DCA11_HUMAN	0.31	0%	No	Possible false positive
5	sp Q05B30 WDR91_XENTR	0.31	0%	No	Possible false positive
6	sp Q8BX09 RBBP5_MOUSE	0.31	1%	No	Possible false positive
7	sp Q24246 DYIN_DROME	0.32	1%	No	Possible false positive
8	sp Q9WUC8 PLRG1_RAT	0.32	1%	No	Possible false positive
9	sp P0CS37 HAT2_CRYNB	0.33	2%	No	Possible false positive
10	sp O74340 DCA13_SCHPO	0.33	1%	No	Possible false positive

Query Protein: A0A010Q8R4

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
--------	---------------	-----	-------------------

IVF-Flat	11.479	0	0.05
----------	--------	---	------

BLAST (Ref)	0.595	1	1.00 (ref)
-------------	-------	---	------------

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: IVF-Flat

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
------	-------------	---------	----------------	-----------------	-------------

1	sp Q6CXX3 HIR1_KLULA	0.28	0%	Yes	-
2	sp A2RAG5 ATG18_ASPNC	0.29	1%	No	Possible false positive
3	sp B8MJJ8 RTC5_TALSN	0.29	0%	No	Possible false positive
4	sp Q24246 DYIN_DROME	0.29	1%	No	Possible false positive
5	sp C0S4K1 ATP25_PARBP	0.29	1%	No	Possible false positive
6	sp E4UPA0 RTC5_ARTGP	0.30	1%	No	Possible false positive
7	sp P0CP30 NPL4_CRYNJ	0.30	0%	No	Possible false positive
8	sp Q8BX09 RBBP5_MOUSE	0.30	1%	Yes	-
9	sp Q4WWM8 SEC2_ASPFU	0.30	1%	No	Possible false positive
10	sp A2QNQ8 GRC3_ASPNC	0.31	1%	No	Possible false positive