

ESG: Extended Similarity Group method for protein function prediction

For any questions regarding ESG contact the administration at info@kiharalab.org

This method allows user to submit protein sequence(s) and obtain annotation prediction for the sequence(s) in terms of [Gene Ontology](#) terms. The underlying ESG algorithm can be found [here](#) . To learn more about each input you can click on the question mark next to it.

Part1: Steps for submitting job as anonymous user

1. Go to <http://www.kiharalab.org/web/software.php> . To start a new job click on the ESG logo highlighted in red below. This will open ESG job submission page in your browser.

Our Software



ESG is our new sequence similarity-based protein function prediction server. In essence, it further applies PFP iteratively and obtains superior performance in terms of prediction accuracy. [Visit the server](#) to submit a sequence or [read the paper in Bioinformatics](#). ESG annotates query sequences with Gene Ontology terms by assigning probability to each annotation. Statistical framework of ESG improves the prediction accuracy by iteratively taking into account the neighborhood of query protein in the similarity based sequence space.



PFP is our sequence similarity-based protein function prediction server designed to predict GO annotations for a query protein sequence beyond what can be found by searching conventional databases. [Visit the server](#) to submit a sequence or [read the paper on Protein Science](#). PFP has achieved the highest total score among participating servers in a function prediction contest held at AFP-SIG'05, ISMB 2005, and also was the best method in the head-to-head performance comparison in the function prediction category at CASP7.

3D-SURFER

3D-Surfer is web-based software for protein surface comparison and analysis. The server integrates repertoire of methods to assist in high throughput screening and visualization of protein surface comparisons. The surface representation enables a very fast structure search; It takes a couple of seconds to perform an exhaustive comparison between a single protein surface to all protein structures in

ESG: Extended Similarity Group Job Submission

Enter Query Sequence(s)

Enter your protein sequence here: [?](#) [Clear](#) [Load Sample](#)

or

Upload your FASTA File: [?](#)

No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#)

Enter the number of stages [?](#)

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

2. You can enter one or more query protein sequence(s) in [fasta](#) format in the text box titled “Enter Query Sequence(s)”.

Consider the following sequence that you can enter.

```
>sp|P56851|EP3B_HUMAN  
MASSLKIWGTLALLCILTLLVQSKEVSWREFMKQHLYLSPSREFREYKCDVLMRENEAL  
KDKSSHMFYIISWYKIEHICTSDNWMDRFRNAYVWVQNPLKVLKCHQENSKNSYTESRSF  
NYIEFHCSMDGYVDSIEDLKMVEPIGN
```

You can also click on “Load Sample” link to load this sequence in the text box and follow the next steps. Clicking on “Clear” link will clear the text box for sequence and load default ESG parameters in the boxes below.

ESG: Extended Similarity Group Job Submission

Enter Query Sequence(s)

Enter your protein sequence here: [?](#) [Clear Load Sample](#)

```
>sp|P56851|EP3B_HUMAN
MASSLKIWGTLALLCILCTLLVQSKEVSWREFMKQHLYLSPSREFREYKCDVLMRENEAL
KDKSSHMF IYISWYKIEHICTSDNWMDRFRNAYVWVQNP LKVLKCHQENSKNSYTESRSF
NYIEFHCSMDGYVDSIEDLKMVEP IGN
```

or

Upload your FASTA File: [?](#)

No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#)

Enter the number of stages [?](#)

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

Or you can click on browse/choose file button and upload a fasta file containing query sequences. Currently we support maximum 1000 query sequences in one job.

To clear sequence, click on clear link next to the label for protein sequence text box.

3. Choose ESG parameters

“Number of hits” indicates number of PSI-BLAST hits to be considered at each level of ESG. Default value of this parameter is 10. “Number of stages” indicates the levels of neighborhood around the query protein that will be considered by ESG. Default value for this parameter is 2. Due to computational constraints currently we support parameters where $\{\text{Number of hits}^{\text{Number of stages}-1}\}$ does not exceed 100. Details about the parameters can be found in ESG documentation over [here](#).

ESG: Extended Similarity Group Job Submission

Enter Query Sequence(s)

Enter your protein sequence here: [?](#) [Clear](#) [Load Sample](#)

```
>sp|P56851|EP3B_HUMAN
MASSLKIWGTLALLCILCTLLVQSKEVSWREFMKQHYLSPSREFREYKCDVLMRENEAL
KDKSSHMF IYISWYKIEHICTSDNWMDFRINA YVWVQNPLKVLKCHQENSKNSYTESRSF
NYIEFHCSMDGYVDSIEDLKMVEPIGN
```

or

Upload your FASTA File: [?](#)

No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#) → Number of hits

Enter the number of stages [?](#) → Number of stages

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

4. When you are submitting the job without logging in, you can book mark the results url and refer back to the job results. If you want to get email notifications about your jobs, view their progress/status and refer to all your jobs at once place, you should register with the website. This will allow you to log in and submit jobs and enjoy additional features free of cost.

5. After loading the query sequence(s) click on submit button. This will take you to the results page and show your job waiting for processing. Bookmark this url and refer to it for the results. Or you can save/remember the job id number to come back to the results.

Job Status

Job ID	Type	Status	Date Submitted
000425	ESG	Waiting for CPU time	2009-12-31

You may bookmark this page to view your results at a later date or use the Job Retrieval form with your Job ID#.

When your job is complete, predicted Gene Ontology terms for the query protein will be displayed. Or you can refresh the page to go to results. The results page will show ESG input parameters, query sequence and prediction terms for each Gene Ontology category (Molecular function, Biological Process and Cellular Component) that have probability greater than 5%. The results page will also provide link to the results in xml format which you can download for further processing.

ESG Job Results

ESG Parameters

Hits per stage: Number of stages:

Protein Sequence:

```
MASSLKIWGTL LALLCILCTLLVQSKEVSWREFMKQHLYLSPSREFREYKCDVLMRENEALKDKSS  
HMFYISWYKIEHICTSDNWMDFRNAYVWQNP LKVLKCHQENSKNSYTE SRSFNYIEFHC SMD  
GYVDSIEDLKMVEPIGN
```

[Download](#)

Molecular Function Terms

Probability	Term	Description
99%	GO:0003676	nucleic acid binding
94%	GO:0004522	pancreatic ribonuclease activity
41%	GO:0003674	molecular_function
15%	GO:0004518	nuclease activity
15%	GO:0004519	endonuclease activity
15%	GO:0016787	hydrolase activity
10%	GO:0005515	protein binding
10%	GO:0003779	actin binding
10%	GO:0004540	ribonuclease activity
10%	GO:0005102	receptor binding
10%	GO:0005507	copper ion binding

Biological Process Terms

Probability	Term	Description
41%	GO:0007286	spermatid development
37%	GO:0007321	sperm displacement
10%	GO:0007275	multicellular organismal development
10%	GO:0030154	cell differentiation
10%	GO:0001525	angiogenesis
10%	GO:0017148	negative regulation of translation
10%	GO:0001666	response to hypoxia
10%	GO:0001938	positive regulation of endothelial cell proliferation
10%	GO:0006651	diacylglycerol biosynthetic process

Cellular Component Terms

Probability	Term	Description
93%	GO:0005576	extracellular region
47%	GO:0005615	extracellular space
16%	GO:0005634	nucleus
10%	GO:0005605	basal lamina
10%	GO:0005730	nucleolus
10%	GO:0032311	angiogenin-PRI complex
5%	GO:0005622	intracellular
7 Predictions		

6. If you close your browser and need to go back to your job results then you can use the “Job Retrieval” text box on the right side of the web page. It has been highlighted in red in the screenshot below. When you enter a valid job id and click on “Fetch Job”, you will get to view the job status or results page depending on if the job is complete or not.

Member Portal Access

Username:

Password:

Login

[Forgot your password?](#)

[Register](#)

Job Retrieval

Job ID#:

Fetch Job

Part2: Steps for logging in and submitting ESG job

1. Open ESG job submission page in your browser.

ESG: Extended Similarity Group Job Submission

Enter Query Sequence(s)

Enter your protein sequence here: [\[?\] Clear Load Sample](#)

or

Upload your FASTA File: [\[?\] Choose File](#) No file chosen

Choose ESG Parameters

Enter the number of hits per stage [\[?\] 10](#)

Enter the number of stages [\[?\] 2](#)

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

Submit

Kihara Lab Software

Thank you for using Kihara Lab Web Servers. Please refer to the tutorials below for getting started with our servers. You can register with the lab and enjoy additional features free of cost. We would be delighted to receive [feedback](#) from you.

Tutorials

[ESG Tutorial](#)
[PFP Tutorial](#)

Member Portal Access

Username:

Password:

[Register](#)

Job Retrieval

Job ID#:

2. Click on the “Register” link on the right side of the page, it is highlighted in red in the screenshot above. It will take you to the account creation page shown below.

New Account Creation

Account Settings

Username: *

Email Address:

Receive Email Notifications

Password: *

Confirm Password: *

* denotes required field

Here you should enter your username and password information. The email address is optional; you should enter it if you wish to receive email notifications about your job status. If you don't check the "Receive Email Notification" check box you will have to log in to the server and check the status of your jobs.

3. Once you have your account set up you can log in and submit jobs for processing. Enter your username and password in the text boxes provided on the right side of the web page and click login. This will redirect you to the software page with links to your jobs appears on right side frame on the page.

Member Portal

[Change Settings](#)
[My Jobs](#)
[logout](#)

Job Retrieval

Job ID#:

You can click on "My Jobs" and view your existing jobs and their status. The complete jobs have clickable links where you can view results of those jobs.

Your jobs

Job ID	Type	Status	Date Submitted	Batch Job
000424	ESG	Done - View or Download	2009-12-31	No
000423	PFP	Done - View or Download	2009-12-31	No
2 Jobs				

To start a new job click on “Software” tab at the top and select ESG by clicking on the logo highlighted in red below.

Our Software



ESG is our new sequence similarity-based protein function prediction server. In essence, it further applies PFP iteratively and obtains superior performance in terms of prediction accuracy. [Visit the server](#) to submit a sequence or [read the paper in Bioinformatics](#). ESG annotates query sequences with Gene Ontology terms by assigning probability to each annotation. Statistical framework of ESG improves the prediction accuracy by iteratively taking into account the neighborhood of query protein in the similarity based sequence space.



PFP is our sequence similarity-based protein function prediction server designed to predict GO annotations for a query protein sequence beyond what can be found by searching conventional databases. [Visit the server](#) to submit a sequence or [read the paper on Protein Science](#). PFP has achieved the highest total score among participating servers in a function prediction contest held at AFP-SIG'05, ISMB 2005, and also was the best method in the head-to-head performance comparison in the function prediction category at CASP7.



3D-Surfer is web-based software for protein surface comparison and analysis. The server integrates repertoire of methods to assist in high throughput screening and visualization of protein surface comparisons. The surface representation enables a very fast structure search; it takes a couple of seconds to perform an exhaustive comparison between a single protein surface to all protein structures in

This will direct you to the ESG job submission page shown in step 1.

4. You can enter one or more query protein sequence(s) in [fasta](#) format in the text box titled “Enter Query Sequence(s)”.

Consider the following sequence that you can enter.

```
>sp|P56851|EP3B_HUMAN
MASSLKIWGTLALLCILCTLLVQSKEVSWREFMKQHLYLSPSREFREYKCDVLMRENEAL
KDKSSHMFIIYISWKIEHICTSDNWMDFRFRNAYVWVQNPLKVLKCHQENSKNSYTESRSF
NYIEFHCSMDGYVDSIEDLKMVEPIGN
```

You can also click on “Load Sample” link to load this sequence in the text box and follow the next steps. Clicking on “Clear” link will clear the text box for sequence and load default ESG parameters in the boxes below.

ESG: Extended Similarity Group Job Submission

Enter Query Sequence(s)

Enter your protein sequence here: [?](#) [Clear](#) [Load Sample](#)

```
>sp|P56851|EP3B_HUMAN
MASSLKIWGTLALLC ILC TLLVQSKEVSWREFMKQHLYLSPSREFREYKCDVLMRENEAL
KDKSSHMF IYISWYKIEHICTSDNWMDFRNAYVWVQNP LKVLKCHQENSKNSYTESRSF
NYIEFHCSMDGYVDSIEDLKMVEPIGN
```

or

Upload your FASTA File: [?](#)

No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#)

Enter the number of stages [?](#)

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

Or you can click on browse button and upload a fasta file containing query sequences. Currently we support maximum 1000 query sequences in one job.

5. Choose ESG parameters

“Number of hits” indicates number of PSI-BLAST hits to be considered at each level of ESG. Default value of this parameter is 10. “Number of stages” indicates the levels of neighborhood around the query protein that will be considered by ESG. Default value for this parameter is 2. Due to computational constraints currently we support parameters where $\{\text{Number of hits}^{\text{Number of stages}-1}\}$ does not exceed 100. Details about the parameters can be found in ESG documentation over [here](#).

6. After loading the query sequence(s) and parameters click on submit button. This will take you to the results page and show your job waiting for processing. You can log in again later to check the status of your job and access the results when it is complete. You can also enter job id in “Job Retrieval” text box on right side of the page and click “Fetch Job” to view the status of job and results. If you have chosen to get email notifications in your account settings then you will be notified by an email when your job is complete.

7. When you job status shows complete, click on the jobID to view predicted Gene Ontology terms for the query protein. The results page will also provide link to the results in xml format which you can download for further processing.

Part3: Steps for submitting a batch job using ESG

1. Open ESG job submission page in your browser.

ESG: Extended Similarity Group Job Submission

Enter Query Sequence(s)

Enter your protein sequence here: [?](#) [Clear](#) [Load Sample](#)

or

Upload your FASTA File: [?](#)

No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#)

Enter the number of stages [?](#)

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

2. You can enter more than one query protein sequences in [fasta](#) format in the text box titled “Enter Query Sequence(s)”.

Consider the following sequences that you can enter.

```
>sp|P56851|EP3B_HUMAN
MASSLKIWGTLALLCILTLLVQSKEVSWREFMKQHLYLSPSREFREYKCDVLMRENEAL
KDKSSHMFYIYSWKIEHICTSDNWMDRFRNAYVWVQNPLKVLKCHQENSKNSYTESRSF
NYIEFHCSMDGYVDSIEDLKMVEPIGN
>sp|P13811|ELBH_ECOLX
MNKVKFYVLFALLSSLCAHGAPQSITELCSEYHNTQIYTINDKILSYTESMAGKREMI
ITFKSGATFQVEVPGSQHIDSQKKAIERMKDTRLRITYLTETKIDKLCVWNNKTPNSIAAI
SMEN
>sp|P05928|VPR_HV1BR
MEQAPEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWHLHGLGQHIYETYGDTWAGVEAI
IRILQQLLFIFRIGCRHSRIGVTQQRARRNGASRS
```

ESG: Extended Similarity Group Job Submission

Enter Query Sequence(s)

Enter your protein sequence here: [?](#) [Clear Load Sample](#)

```
>sp|P56851|EP3B_HUMAN
MASSLKIWGTLLALLCILCTLLVQSKEVSWREFMKQHLYLSPSREFREYKCDVLMRENEAL
KDKSSHMF IYISWYKIEHICTSDNWMDRFRNAYVWVQNPLKVLKCHQENSKNSYTESRSF
NYIEFHCSMDGYVDSIEDLKMVEPIGN
>sp|P13811|ELBH_ECOLX
MNVKVFVVLFTALLSSLCAHGAPQSITELCSEYHNTQIYTTINDKILSYTESMAGKRENV I
```

or

Upload your FASTA File: [?](#)

No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#)

Enter the number of stages [?](#)

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

Or you can click on browse button and upload a fasta file containing query sequences. Currently we support maximum 1000 query sequences in one job.

3. Follow steps 3 to 6 from part 1 if you are submitting the job as anonymous or steps 5 to 7 from part 2 if you are logged in as a registered user. On the job status page the job type will be shown as batch.

Your jobs

Job ID	Type	Status	Date Submitted	Batch Job
000401	ESG	Job Processing	2009-12-19	Yes
000400	ESG	Done - Results	2009-12-19	No

Once the processing is complete the results will be displayed in the table with each sequence, its identifier and a clickable link which will take you to detailed results for that sequence.