InterPro: Functional and structural analysis of protein sequences

Alex Mitchell [1]

- Proteins
- Beginner
- 1 hour

InterPro is a classification database that provides predictive information about protein sequences. This course will show you what kind of information is available in InterPro, how to access it, and how to search the database.

Learning objectives:

- Be able to search the InterPro database in a number of different ways
- Be able to interpret the information in an InterPro entry page
- Be able to review the graphical representation of the signature matches on an InterPro protein page

What is InterPro?

InterPro [2] is a bioinformatics resource that provides functional analysis of protein sequences by classifying them into families and predicting the presence of domains and important sites (Figure 1). To classify proteins in this way, InterPro uses predictive models, known as signatures, provided by several different databases (referred to as member databases) that make up the InterPro Consortium.
Figure 1. The InterPro homepage can be found here [3] and is easily accessible by clicking on the InterPro link in the Proteins panel on the EBI services [4] page.

Where does the data come from?

In order to classify proteins into families and to predict the presence of important domains or sequence features, we require computational tools. One such set of tools are predictive models known as protein signatures.

Signatures are built by the member databases in the InterPro [2] consortium (Figure 2). Different member databases use different methods to construct their signatures, and they have their own particular focus of interest: structural and/or functional domains, protein families, or protein features such as active sites or binding sites.

To learn more about the different methods that can be used to classify proteins and the different types of models used by InterPro's member databases, see our Introduction to Protein Classification [5] tutorial.
Why do we need InterPro?

Protein signature [6] databases have become vital tools for classification of protein sequences in order to infer their function. Many protein signature-based resources are now available, each with their own strengths and weaknesses.

We need InterPro [2] because it:

- Reduces redundancy and simplifies protein sequence analysis by integrating signatures from different member databases that represent the same protein family [7], domain [8] or site.

- Unites the member databases, capitalising on their individual strengths to produce a powerful classification tool.

- Provides a single convenient searchable location, allowing simultaneous querying of all member databases.

- Adds information (including descriptive abstracts and Gene Ontology [9] terms) to the signatures, which may be used to annotate the proteins they match.
How can InterPro help with your research?

When to use InterPro

You can use InterPro if you have an amino acid sequence or set of sequences and you want to know:

- what they are - the family to which they belong
- what their function is and how it can be explained in structural terms

You can also use InterPro for a variety of other purposes, such as predicting GO terms for a set of sequences, or identifying all sequences in the UniProtKB database that are predicted to belong to a given family or to possess a particular domain.

When not to use InterPro

InterPro cannot help you if:

- you want to perform structural alignment of protein sequences
- you have a genomic DNA sequence and are interested in gene annotation (intron / exon predictions, identification of promoter regions, etc)

What is an InterPro entry?

At its heart, each individual InterPro entry consists of one or more member database signatures that are characteristic of the same protein family, domain or sequence feature. The entry is given a name and a unique InterPro identifier. It is also annotated with additional information, which can be found in different sections on the entry page.

For example, Figure 4 below shows the InterPro entry for the type 2 malate dehydrogenase protein family. The entry has a name (malate dehydrogenase, type 2) and identifier (IPR010945). The contributing signatures are shown on the right-hand side, with links to the signatures on the individual member database web sites. A descriptive abstract explains what these proteins are and what their function is. It also provides links to publications in PubMed for more detailed information. A set of GO terms is also provided, which describe the characteristics of the proteins matched by the entry.
Figure 4. Overview of a typical InterPro entry

InterPro entry types

InterPro [2] entries are classified into one of five categories, depending on the biological entity they represent: homologous [14] superfamily [15], protein family [7], domain [8], repeat or site.

The entry type is indicated by a specific icon (Figure 5), which can be found on the top left hand side of an InterPro entry page [16].
InterPro: Functional and structural analysis of protein sequences
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

Figure 5. Icons denoting the different type of entries (homologous superfamily, family, domain, repeat or site) that can be found in the InterPro database.

InterPro entry types: Family and Homologous Superfamily

An InterPro [2] protein family [7] is a group of proteins that share a common evolutionary origin, reflected by their related functions and similarities in sequence or structure. Protein families are often arranged into hierarchies, with proteins that share a common ancestor [17] subdivided into smaller, more closely related groups. For example, steroid hormone receptors constitute a family of nuclear receptors responsible for signal transduction [18] mediated by steroid hormones, and can be subclassified into different groups, including the liver X receptor subfamily (Figure 6). This subfamily consists of nuclear receptors that regulate the metabolism [19] of several important lipids, including oxysterols.

![Family hierarchy of steroid hormone receptors](image)

Figure 6. Example of a protein family hierarchy. The steroid hormone receptor family can be subdivided into a number of smaller, closely related subfamilies.

A Homologous [14] Superfamily [15] is a large diverse family, usually with a shared tertiary structure. They are exclusively composed of CATH [20]-Gene3D [21] and/or SUPERFAMILY methods, both of which utilise a collection of underlying profile hidden Markov models (HMMs) to represent diverse structural families, rather than one single model. Unlike Family entries, Homologous Superfamily entries do not exist in hierarchies.

InterPro entry types: Domain

Domains are distinct functional and/or structural units in a protein. Usually they are responsible for a particular function or interaction, contributing to the overall role of a protein. Domains may exist in a variety of biological contexts, where similar domains can be found in proteins with different functions.

![Domain representation](image)

For example, the pleckstrin homology (PH) domain [8] is a small modular domain that occurs in a large variety of proteins and is involved in phospholipid binding. One group of proteins containing a PH domain are the beta-adrenergic receptor kinases (Figure 7). Four domains have been
identified in these proteins: an RSG (regulator of G protein signalling) domain, a protein kinase (PK) domain, an AGCK domain, involved in regulation by phosphorylation, and a C-terminal PH domain.

![Domain Architecture](image)

**Figure 7.** Graphical representation of the domain architecture of beta-adrenergic receptor kinases.

**InterPro entry types: Sites and Repeats**

**Sites**
Sites are groups of amino acids that confer certain characteristics upon a protein, and may be important for its overall function. Sites are usually quite small (often only a few amino acids long). The types of site covered by InterPro [2] are:

- active sites, which contain amino acids involved in catalytic activity
- binding sites, containing amino acids that are directly involved in binding molecules or ions
- post-translational modification (PTM) sites, which contain residues known to be chemically modified (phosphorylated, palmitoylated, acetylated, etc) after the process of protein translation [22]
- conserved sites, which are found in specific types of proteins, but whose function is unknown.

For example, the stretch of residues involved in the catalytic function of the S1B subfamily of serine peptidases is specific to this type of peptidases and constitutes the *active site* [23] of these proteins.

**Repeats**
Repeats are typically short amino acid [10] sequences that are repeated within a protein, and may confer binding or structural properties upon it.

For example, pentapeptide repeats are sequence motifs of five amino acids found in multiple tandem copies. They were first identified in cyanobacterial proteins, where they can be found in many copies. Their function is currently unknown.
Relationships between InterPro entries

InterPro [2] entries often share relationships with other entries in the database. For example, an entry might represent a functionally-specific subfamily of a larger family, or a subclass belonging to a broad class of domains. These relationships are stored in InterPro as hierarchies, which are displayed as trees in the relationship section of an entry page (Figure 8).

Both family and domain [8] entries are able to form hierarchical relationships, although family and domain hierarchies are kept separate in the database and do not overlap (for example, a subclass of a particular domain cannot be a subtype of a protein family [7]). However, an exception is made for the relationship between homologous superfamily with family and domain entries; it is generated automatically using the Jaccard and containment indexes. This relation appears in the Overlapping homologous superfamilies section (see figure). Repeats and sites are not arranged into hierarchical relationships.

Figure 8. Examples of hierarchical relationships between InterPro entries

For more information on the Homologous superfamily entry type in InterPro, have a look at our Genome3D annotations in InterPro tutorial [24].

Top menu links

More information is available from the top menu of an InterPro [2] entry page, including links to the proteins matched by that entry and the species in which they are found.

The different types of information found in this menu are shown in Figure 9.

A more detailed description of some of the information available from the top menu is given in the following sections, where we explore searching the database.
**Figure 9.** Overview of the information available from the top menu of an InterPro entry

**How to search InterPro?**

You can search InterPro [2] (Figure 10) using either an amino acid [10] sequence, a text search (e.g., with a word or short phrase, or a UniProt [25] or InterPro identifier [12]), a domain [8] architecture or browse through the InterPro member databases and/or entry types. The following sections will take you through these options.
Figure 10. Search fields on the InterPro home page, showing text search field (A) and the sequence search (B) options, including 'Advanced search', where you can limit your search to those member databases or sequence features of interest. Selecting the browse tab in the top menu (C) allows access to a browse search, (e.g., search for member database signature, InterPro entry type). You can also search for a particular domain architecture (D).

Sequence searching

To search InterPro [2] with a protein sequence, copy and paste an amino acid [10] sequence into the large box, or upload a fasta [26] file on the home page and click on the search button located below the box (Figure 11). InterPro will then attempt to assign the sequence to a protein family [7] and identify any domains, repeats and sites. The Advanced options allow you to select a particular member database, a sequence feature or to change the JOB ID.
Figure 11. InterPro sequence search box

When clicking on the search button, you are redirected to the InterProScan [27] searches page (Figure 12).

Figure 12. InterProScan searches page

Once the analysis is completed, clicking on the JOB ID, will lead you to a results page showing the InterPro matches to your query sequence in graphical form (Figure 13). Next we will look at interpreting these results to understand what they mean, using a guided example.
Figure 13. Results page showing analysis of an amino acid sequence using InterPro

Guided example: searching InterPro with an amino acid sequence

Getting started

In this guided example, we will look at searching InterPro [2] with an amino acid [10] sequence.

First, we need to locate the sequence search box on the InterPro homepage.
Navigate to the InterPro [3] homepage and find the sequence search box.

Now we need a sequence to search against the database.

Copy and paste the following sequence into the sequence search box, and press 'Search'

>MY_SEQ
MAAGGSGAESAPTPSMSSLPLAALNVRHRSLFLNVRTQVAADWTGLAEEM
NFYELEIRLETHYPETRSLDDLQGRPASVGRLLLLAKLGRDDVLLGLPS
IEEDCRKYILKQQEAAEKLQVDVDSSSIPWMSGITIDPDLQMPEDFADI
CYCPSDIQFVQEMIRQLEQTNRLKLCVSDDRVLPGTCVWSIASELIEKRCRRM
VVVVSDDYLSKESCDFQTKFALSLSPGAHQKRLIPVKYKMKKFSILRFITV
CDYTNPCPKSWFWRALSLP

We will now go through the search results over the next few pages.

**Sequence search results: overview page**

The search result page should look something like the one shown in Figure 14 below. This is an interactive page that contains a large amount of information.
Figure 14. InterPro sequence search results overview page

Section A shows the top menu, allowing to get all the InterPro entries matches the sequence (Figure 15) and the sequence with the possibility to do another search through InterProScan [27] or through the HMMER server (Figure 16).

Section B allows to download the results in different formats.

All the InterPro entries and member databases found are displayed in the Protvista sequence viewer. The searched sequence is represented by a grey bar at the top of the sequence viewer with its length in amino acids displayed along the top. Matching InterPro entries and member database signatures are indicated as coloured bars.

Section C is the interactive menu which can be used to control the sequence viewer display, change the view to full screen mode or take screenshots.

Section D shows InterPro entries found grouped by InterPro entry types. Mousing over the bars reveals the type of domain [8] that they represent, along with their position on the sequence.

Section E summarises the domain InterPro predicts the protein to contain based on the signatures integrated in
the entry.

Section F displays the member database signatures domains.

Section G shows information about residues hydrophobicity.

Section H shows the Gene Ontology [9] (GO) terms predicted for the protein. These terms are assigned based on the matches to the InterPro entries shown above.

Figure 15. InterPro entries matching the searched sequence

Figure 16. Submitted sequence with link to InterProScan and HMMER search

Sequence search results: family information

Looking at the results overview page, we will examine the family that InterPro predicts the query protein to contain.
The InterPro entry pages and information available for all the different InterPro entry types are similar.

On the results page, place your mouse over the InterPro family in the sequence viewer (Figure 17).

Figure 17. Predicted protein family for the query sequence

Mousing over the family, we see its InterPro accession number (IPR017281), short name (Myelin_different_respons_MyD88), InterPro entry type (family) and its position on the sequence (amino acid residues 1-293).

Click on the InterPro accession to visit the entry page for myeloid differentiation primary response protein MyD88.

Clicking on the link will take us to the InterPro entry page for this family (Figure 18). This page explains what MyD88 is, what its functions are, and the GO terms that can be applied to members of the family. Clicking on the top menu allows us to examine the different UniProt proteins that are predicted to belong to the family (Proteins), their different domain architectures (Domain Architectures), the species in which they are found (Taxonomy), the reference proteomes in which they are found (Proteomes), their solved protein structures (Structures) and predicted domains (Genome3D). Clicking on any of these menu items opens a new page with the appropriate information and links that allow you to download specific datasets (e.g., all the proteins in a family, or all the proteins with a particular domain architecture, or all the protein family members from a particular kingdom, class or species).
Figure 18. InterPro entry page for the myeloid differentiation primary response protein MyD88 family of proteins

Sequence search results: exploring other proteins in the family

It is possible to find all of the UniProtKB [11] proteins that match a protein family [7], using the ‘Proteins’ link on the left hand side of the family page (Figure 19).
family can be found.

Click the 'Proteins' link to see other proteins in UniProtKB with the same family membership.

Following the 'Proteins' link takes us to a page like the one shown in Figure 20. This is a paginated list of UniProtKB proteins, showing their name, accession number, species and domain architectures (if known). Those sequences with a tick symbol are from Swiss-Prot, the manually annotated section of UniProtKB, whilst others are from the automatically annotated section TrEMBL. A file in different formats containing the sequences of all of the proteins can be downloaded by clicking the 'Export ' button at the top right of the page.
Searching InterPro with a batch of amino acid sequences

Often, people want to search InterPro with a batch of amino acid sequences - for example, to analyse whole genomes or proteomes. There are two ways in which it is possible to do this. The first way involves downloading and installing the InterProScan [27] analysis tool. This can be obtained from here [30].
The software comes with full installation instructions. It also has a lot of configurable features, including the ability to search nucleic acid sequences against the database.

The second way of searching InterPro with a batch of sequences involves the use of the EBI's web services [31]. These allow up to 30 sequences to be analysed per request. Instructions on how to use the SOAP [32]-based web service are available here [33].

Instructions on using the REST [34]-based web service can be found here [35].

**Text search**

InterPro can also be searched using text (Figure 21). The different search types accepted are listed below:

<table>
<thead>
<tr>
<th>Type of search term</th>
<th>Result of search</th>
</tr>
</thead>
<tbody>
<tr>
<td>InterPro identifier</td>
<td>A list of InterPro entries and member database signatures related to the identifier</td>
</tr>
<tr>
<td>Word or a short phrase</td>
<td>A list of all InterPro entries and member database signatures with the search word or phrase in their title or annotation</td>
</tr>
<tr>
<td>Member database signature</td>
<td>The InterPro entry it is integrated in</td>
</tr>
<tr>
<td>Gene Ontology term identifier</td>
<td>A list of all InterPro entries and member database signatures with the Gene Ontology term</td>
</tr>
<tr>
<td>Structural identifier</td>
<td>A list of InterPro entries and member database signatures related to the structural identifier</td>
</tr>
<tr>
<td>UniProtKB protein identifier</td>
<td>A list of InterPro entries and member database signatures related to the UniprotKB protein identifier</td>
</tr>
<tr>
<td>Proteome identifier</td>
<td>A list of InterPro entries matching the proteome</td>
</tr>
<tr>
<td>Set identifier</td>
<td>A list of member database signatures included in the set.</td>
</tr>
</tbody>
</table>

For all the term searched, except for the word/phrase and Gene Ontology search, a link to the corresponding InterPro page is available in the 'found an exact match' section in the search result page if included in InterPro.

When entering an InterPro identifier, a UniProtKB identifier, a member database signature, a protein structure identifier, a proteome identifier or a set identifier directly in the top search (see figure) from any InterPro page, you will be redirected to their respective InterPro page; this is explained further in the next section.

![Figure 21. To search InterPro using a word, phrase or identifier type it in the text search box as indicated](image)

**Searching with a protein identifier**
By searching for a protein identifier on the quick search or by clicking on it in the search result list, you'll be redirected to the InterPro protein page.

The protein page is similar to the sequence search results page, but it contains extra information, drawn from UniProtKB which is displayed at the top of the page in an additional section (Figure 22). This includes information on the name of the protein in UniProtKB, its accession number and identifier, the species in which it is found and its length in amino acids. Following the external link section will take you to the UniProtKB entry for this protein.

The page also contains a ‘family membership’ section showing the InterPro entries the protein is found in.

Additional information can also be found in the top menu: ‘Entries’ lists all the InterPro entries where the protein is found. ‘Similar proteins’ lists all UniProtKB proteins with the same domain architecture, and ‘Structures’ lists any structural information in UniProt for the protein accession searched (in each case, these links will only be active where such information is present in UniProtKB).

In the protein sequence viewer, family, domain, homologous superfamily, site and repeat information, are displayed as normal, but extra information is available: Mobidb consensus disorder and Genome3D domain and structure predictions. You can also generate residue conservation information for regions of the protein sequence matching a Pfam domain.

For more information regarding Genome3D predictions, the Genome3D annotations in InterPro course will be available on trainonline soon.
Figure 22. Protein page showing InterPro predictions for UniProtKB protein O88551

Searching with a protein structure identifier

By searching for a protein structure identifier [12] with the quick search or by clicking on it in the search results list,
you'll be redirected to the InterPro [2] structure page.

The structure page, like the protein page, is also similar to the sequence search result page, but with extra information as shown in Figure 23. This contains information on the name of the structure in PDBe [36], its accession [28] identifier, its resolution, chains and release date. On the right side, a link to the structure page in the Protein Data Bank in Europe is available.

On the top menu, 'Entries' and 'Proteins' list all the InterPro entries and UniProtKB [11] proteins where the structure is found.

There is also a 3D structure viewer, where you can interact with the structure (e.g. turn, zoom...) and select a member database signature or an InterPro entry to visualise which part of the structure it matches.

For each chain in the structure, a protein sequence viewer shows the secondary structure [37] information, Genome3D [38] domain [8] predictions, and the InterPro entries and member database signatures overlapping the chain.
Figure 23. InterPro structure page

Searching with a member database signature
By searching for a member database signature on the quick search, then clicking on the signature [identifier][12] in the search result list or on the protein sequence viewer, you'll be redirected to the [InterPro][2] member database signature page.

The member database signature page (Figure 24) contains information about the member database providing the signature, the signature type, its short name, full name, a description and references if available. On the right, a link to the member database website is available under the 'External links' section. If the signature is integrated in an InterPro entry, the entry is displayed under the 'Integrated to' section.

Figure 24. InterPro PRINTS signature page for PR01377

The top menu varies depending on the data available for each signature. It contains at least the 'Proteins' (proteins matching the signature), 'Taxonomy' (list of species the signature matches) and 'Proteomes' (reference proteomes) tabs.

Additional links that are displayed only when the data is available include 'Structures' ([PDB][39] structures matching the signature), 'Sets' (set of signatures it belongs to) and 'Signature' (residue conservation hmm logo), as shown in Figure 25.

Figure 25. InterPro Pfam signature page for PF00822
Browse search

If you aren't interested in searching for a particular protein sequence or text, the browse search (Figure 26) allows you to search through InterPro for a list of InterPro entries, proteins, structures, species, proteomes or sets corresponding to selected criteria. While the search is processing, the InterPro logo is moving.

Figure 26. Browse search example. Result of the UniprotKB proteins found in the PRINTS database for the *Saccharomyces cerevisiae* organism

Summary

- InterPro is a resource for classifying proteins into families and predicting domains, repeats and functional sites.
- InterPro integrates protein signatures from 13 member databases that are part of the InterPro consortium: CATH-Gene3D, CDD, HAMAP, PANTHER, Pfam, PIRSF, PRINTS, PROSITE Patterns, PROSITE Profiles,
SFLD, SMART, SUPERFAMILY and TIGRFAMs. Protein signatures are predictive models based on similarities among proteins that have the same structure or function.

- You can search InterPro with an amino acid sequence using the web site.

- The InterPro web site can also be searched with a word or phrase, an entry ID, UniProtKB accessions and identifiers, GO terms and structural identifiers.

- The InterPro web site integrates a browse search functionality.

- Batches of sequences can be searched against InterPro by downloading the InterProScan tool, or using the EBI's web services.

- The InterPro protein view provides a graphical representation of the signatures that match a particular protein, with information about protein family membership, sequence features, structural features, and structural predictions for that protein.

- The InterPro structure view provides a graphical representation of the signatures that match a particular protein structure with a 3D structure viewer, information about the chains secondary structures, InterPro entries and proteins membership.

### Quiz: Interpro Quiz

**Questions:** 5

**Attempts allowed:** Unlimited

**Available:** Always

**Pass rate:** 75%

**Backwards navigation:** Allowed

### Exercises

This exercise allows you to apply the knowledge you have gained on this course by providing an example of how InterPro [2] can be used and asking you to solve a given task. Start by clicking on the exercise title provided.

If you need help to complete this section you can look in the 'Need some help?' and 'Want to know how we did it?' sections.

### Analyse a protein sequence

**Scenario**

You have isolated a protein from squirrel. You want to know more about this protein, perhaps the presence of
certain domains or features can give you a clue about its role.

**Exercise**

Find information about the following protein using InterPro.

```plaintext
>squirrel_seq
MALPARLVPLCCLALLALPAOQSCGPGRPVGRRRYVRKQLVPLLYKQ
FVPSVPERTLGASGPAGRVARSGERFREDLVPWNYPDIIIFKDEENSG
ADRLMTERCKERVNAIAAVNMWPGVRLVTEGDGDEGDQHQAQDSLH
YEGRALDITTSDRRNKYLALLARSEAGFDVYRYYESRNHABAVEVKA
GTVGGCFRETEAAQLWDVAARGRELHRAWLAADAAGRVRVPTVLL
FLDRDLQRRASFVAVETERPPRLTPWHVFAARGPAPAPGDFAP
VFARRLRAGDSVLAPGDDALRPVARVAREEAAVGVPAPLTAHGTLL
VNDVLASCYAVLESQAHRAFAPLRLHALGALLPGGAVQPTGMHW
YSRFLYRLAEELLG
```

**Need some help?**

1. Copy and paste the FASTA sequence into the 'analyse your protein sequence' box on the InterPro home page and press 'Search'.
2. Once you are in the protein overview page, explore the information on the graphical results view. What family does this protein belong to? What domains does it have? What processes is it involved in?
3. Investigate further the entries that match this protein by clicking in the InterPro identifiers and reading the descriptions for those entries.

**Want to know how we did it?**

When you search InterPro with the FASTA sequence provided, you will get a graphical representation of results, similar to that shown in Figure 27.
From the results in InterPro you can infer:

- That the protein is predicted to contain an N-terminal signalling domain, which is part of a wider class of domains that also includes the zinc-binding domains of DD-peptidase.
- That the protein is also predicted to contain a C-terminal Hint domain, which is involved in autoprocessing of the protein. The Hint domain itself can be broken down into two smaller subdomains.
- That the protein queried is predicted to a member of the hedgehog protein family, and therefore involved in a range of biological processes, including development and cell-cell signalling. How? From the GO terms and description from entry IPR001657 (see Figure 28 below).
Figure 28. The InterPro entry page for IPR001657 (Hedgehog protein).

Your feedback

Please tell us what you thought about this course. Your feedback is invaluable and helps us to improve our courses and thus enhance your learning experience.

Get help and support on InterPro
Documentation

You can find documentation about InterPro [2] (the user manual, release notes, etc) linked from our web page [3] or you can download the information from our FTP server [40].

Contact

For all support enquiries, please contact the InterPro interhelp [at] ebi.ac.uk (help desk).

Learn more

Recommended courses

Online courses:

Introduction to Protein Classification at the EBI [5]

Genome3D annotations in InterPro [41]

The EBI offers hands-on courses covering a range of subjects, including training on InterPro. An up-to-date list of training courses is available here [42].

Contributors

Alex Mitchell [1]

EMBL-EBI
Curation co-ordinator, InterPro and EBI Metagenomics

Alex Mitchell is co-ordinator for the InterPro and EBI Metagenomics databases at EMBL-EBI. He obtained his DPhil in pharmacology from the University of Oxford, and was previously employed as a molecular biologist at the Institute of Psychiatry. He moved to the University of Manchester to work on protein sequence analysis and functional classification, before joining EMBL-EBI in 2011.
Amaia Sangrador [43]

EMBL-EBI
Scientific curator, InterPro

Amaia Sangrador is a curator for InterPro at the European Bioinformatics Institute in Cambridge, UK. She joined EMBL-EBI in 2010. She has a PhD in Molecular Biology and has been involved in several research projects in the areas of immunology and comparative genomics.

Mindi Sehra [44]

EMBL-EBI
EMBL-EBI ALUMNI Scientific Training Officer - eLearning Content Developer

Mindi Sehra was the Scientific Training Officer (eLearning) for the Outreach and Training Team at EMBL-EBI. Mindi was responsible for expanding and consolidating the EBI’s range of online training materials and monitoring and maintaining the portal, which includes investigating ways to exploit electronic technologies. Mindi completed a Genetics Degree at Sheffield University before moving into genome analysis at the Wellcome Trust Sanger Institute. She completed a MSc in Medical Genetics and Immunology at Brunel University, her thesis on the Swine Leukocyte Antigen secured her a position in the Human and Vertebrate Annotation [45] and Analysis group as a computer biologist. She then joined the UniProt [25] team at the EBI as a protein curator [46] working on automatic and manual annotation [47].
Typhaine Paysan-Lafosse [48]
EMBL-EBI
Bioinformatician

Typhaine Paysan-Lafosse is working as a bioinformatician and scientific curator [46] for the InterPro [2] database, previously she worked on the Genome3D project for InterPro and PDBe [36] databases. She joined EMBL-EBI in 2016. Beforehand she worked for the international ImMunoGeneTics information system in Montpellier.

References


Source URL: https://www.ebi.ac.uk/training/online/course/interpro-functional-and-structural-analysis-protein-sequences

Links
[1] https://www.ebi.ac.uk/training/online/trainers/mitchell
[2] https://www.ebi.ac.uk/training/online/glossary/interpro
[5] https://www.ebi.ac.uk/training/online/course/protein-classification-introduction-embl-ebi-resou
[6] https://www.ebi.ac.uk/training/online/glossary/protein-signature
[7] https://www.ebi.ac.uk/training/online/glossary/protein-family
[8] https://www.ebi.ac.uk/training/online/glossary/domain
[9] https://www.ebi.ac.uk/training/online/glossary/gene-ontology
[10] https://www.ebi.ac.uk/training/online/glossary/amino-acid
[11] https://www.ebi.ac.uk/training/online/glossary/uniprotkb
[12] https://www.ebi.ac.uk/training/online/glossary/identifier
[13] https://www.ebi.ac.uk/training/online/glossary/pubmed
[14] https://www.ebi.ac.uk/training/online/glossary/homologous
[15] https://www.ebi.ac.uk/training/online/glossary/superfamily
[16] https://www.ebi.ac.uk/training/online/course/interpro-functional-and-structural-analysis-protei
[17] what-interpro-entry
[18] https://www.ebi.ac.uk/training/online/glossary/ancestor
[19] https://www.ebi.ac.uk/training/online/glossary/signal-transduction
[20] https://www.ebi.ac.uk/training/online/glossary/metabolism
[21] https://www.ebi.ac.uk/training/online/glossary/cath
[22] https://www.ebi.ac.uk/training/online/glossary/gene3d
[23] https://www.ebi.ac.uk/training/online/glossary/active-site
[24] https://www.ebi.ac.uk/training/online/course/genome3d-annotations-interpro/homologous-superfamily-entry-type-interpro
[25] https://www.ebi.ac.uk/training/online/glossary/uniprot