

# Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users

Séverine Duvaud<sup>1,\*</sup>, Chiara Gabella<sup>1,\*</sup>, Frédérique Lisacek<sup>2,3</sup>, Heinz Stockinger<sup>1</sup>, Vassilios Ioannidis<sup>1</sup> and Christine Durinx<sup>1</sup>

<sup>1</sup>SIB Swiss Institute of Bioinformatics, Quartier Sorge - Bâtiment Amphipôle, CH-1015 Lausanne, Switzerland,

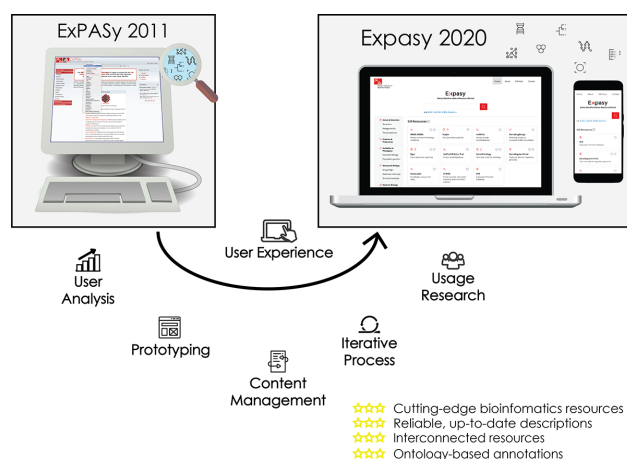
<sup>2</sup>Proteome Informatics Group, SIB Swiss Institute of Bioinformatics, and Computer Science Department, University of Geneva, CH-1227 Geneva, Switzerland and <sup>3</sup>Section of Biology, University of Geneva, CH-1205 Geneva, Switzerland

Received February 02, 2021; Revised March 11, 2021; Editorial Decision March 13, 2021; Accepted April 01, 2021

## ABSTRACT

The SIB Swiss Institute of Bioinformatics (<https://www.sib.swiss>) creates, maintains and disseminates a portfolio of reliable and state-of-the-art bioinformatics services and resources for the storage, analysis and interpretation of biological data. Through Expasy (<https://www.expasy.org>), the Swiss Bioinformatics Resource Portal, the scientific community worldwide, freely accesses more than 160 SIB resources supporting a wide range of life science and biomedical research areas. In 2020, Expasy was redesigned through a user-centric approach, known as User-Centred Design (UCD), whose aim is to create user interfaces that are easy-to-use, efficient and targeting the intended community. This approach, widely used in other fields such as marketing, e-commerce, and design of mobile applications, is still scarcely explored in bioinformatics. In total, around 50 people were actively involved, including internal stakeholders and end-users. In addition to an optimised interface that meets users' needs and expectations, the new version of Expasy provides an up-to-date and accurate description of high-quality resources based on a standardised ontology, allowing to connect functionally-related resources.

## GRAPHICAL ABSTRACT






## INTRODUCTION

Databases, software tools and on-line services are essential resources in the daily work of most life scientists. They are key to the long-term preservation of scientific data and the reproducibility of life science studies. As on-line bioinformatics usage spread and diversified, portals offering single access points were created (1,2) and later on, structured catalogues were proposed (3,4). In this landscape, the Expasy portal pioneered. It was created in 1993 referred to as 'the Expert Protein Analysis System' with a primary focus on protein knowledge (5). It was the first life science website - and among the 150 very first websites in the world. In the past 27 years, the portal has evolved. It was redesigned several times and the penultimate version was released in 2011, when it became the 'Expasy SIB Bioinformatics Resource Portal', a catalogue of bioinformatics resources (6). As detailed in (7) SIB Swiss Institute of Bioinformatics federates a Swiss bioinformatics community now close to 800 scientists. In this context, hundreds

\*To whom correspondence should be addressed. Tel: +41 21 692 40 50; Email: [severine.duvaud@sib.swiss](mailto:severine.duvaud@sib.swiss)  
Correspondence may also be addressed to Chiara Gabella. Tel: +41 21 692 40 50; Email: [chiara.gabella@sib.swiss](mailto:chiara.gabella@sib.swiss)  
†The authors wish to be known that, in their opinion, this first two authors should be regarded as joint First Authors.

**Table 1.** Examples of user-types (personas) and their use cases

Persona	Goal – why does (s)he need Expasy?
 <p>Miranda, Jr Biologist</p>	Just starting a PhD, she wants to have a clear vision of what life science areas are covered in bioinformatics and what she can do with the various databases and software tools.
 <p>Chloe, Bioinformatician embedded in a research lab</p>	Looking for a database able to provide expression data in a tab delimited format.
 <p>Kris, Sr Biologist</p>	Looking for a general overview of information on the MEF2C gene. Wants to quickly gather all data available in the various databases.

of cutting-edge resources are developed and made available to the life science community. Among those, twelve key resources—the so-called SIB resources—receive specific support from the Institute and form the SIB Resource portfolio that ranges from specialised knowledge bases such as UniProtKB/Swiss-Prot (8) (part of the UniProt consortium), neXtProt (9), STRING (10) and Bgee (11), to online tools such as SWISS-MODEL (12) and SwissDrugDesign (13).

In this article, we describe the latest version of Expasy together with its redesign process that relied on rules of User-Centred Design (UCD). UCD combines User Research (UR), which addresses the question ‘Who is the target audience?’, with User Experience (UX), which addresses the question ‘How does the user interact with the application?’. Originating from the computer industry, UCD is emerging in science, as its value has been acknowledged (14,15). Expasy was improved to maintain access to SIB’s high-quality scientific resources and to offer up-to-date and accurate information on each of them, with a more efficient and easier-to-use search engine. Additionally, a networked organisation of resource descriptions was implemented, to broaden the scope of the initial search and explore the diversity of the SIB resources. The last new aspect to the portal is the involvement of resource providers in the content update of Expasy as a crucial step toward guaranteeing quality information. With this approach, the new version focuses specifically on resources developed by or in collaboration with SIB groups, in contrast with previous editions.

## OVERVIEW

### Purpose

Expasy is an extensible and integrative portal including >160 databases and software tools developed by SIB groups. It covers a wide range of fields in life sciences and biomedical research, spanning genomics, proteomics, struc-

tural biology, evolution, phylogeny, systems biology and medicinal chemistry. Expasy is an essential tool for a large variety of users, from beginners interested in discovering bioinformatics, to advanced scientists looking for specific biological answers. Table 1 illustrates three examples of typical users of the portal (so-called personas, see below) and their corresponding use-cases. Given its usefulness for the life science community, Expasy has undergone a major overhaul in 2020. The new user interface is the result of in-depth work with users to meet their needs and expectations. Its content is a long-term commitment of the resource provider community to describe each resource accurately and in a standardised manner.

### Homepage

Figure 1 shows the homepage of the new Expasy portal, including its new logo. A search bar is directly accessible from the top of each page. A filter panel is proposed on the left side of the page. It reflects the categories defined to filter the resources (see definition below). All resources are listed as cards with the resource’s name, a short description, type (database and/or tool) and categories.

### Search and result page

The portal provides a single search bar allowing simultaneously two types of search: a ‘regular search’ and a ‘cross-resource search’. The results are displayed jointly on one page.

*Regular search: browsing the resources.* Expasy allows users to search for resources by name, keyword, category or description. When a user, for example, types ‘virus’ in the search bar, this results in a list of virus-related databases and software tools, such as V-pipe (16), ViralZone (17), OpenFlu (18) or COVID-19 Scenarios (19). We will refer to ‘regular search’ in the rest of the article to designate

The screenshot shows the Expasy homepage. At the top right, there is a navigation menu with 'Home', 'About', 'SIB News', and 'Contact'. Below this is the 'Expasy' logo and 'Swiss Bioinformatics Resource Portal'. A search bar is located at the top center, with a magnifying glass icon and a red search button. Below the search bar, there is a list of 'SIB Resources' and 'Other Resources of SIB Groups'. On the left side, there is a filter panel with various categories and sub-categories, each with a checkbox. The 'UniProtKB/Swiss-Prot' card is highlighted with a yellow box. Three dashed lines with numbers 1, 2, and 3 point to the search bar, the filter panel, and the resource cards, respectively.

**Figure 1.** Elements on the homepage of Expasy: 1. Single search bar at the top; 2. Filter panel on the left and 3. Resources listed as cards and sorted randomly in the main section of the page.

this type of search. All the resource-specific information is stored in a relational database back-end and queried using Elastic Search (<https://www.elastic.co/elasticsearch/>, (20)), with fuzzy search or approximate string-matching search enabled. Resources are linked one-to-the-other through ontology-based terms. As a result, they form a network of seamlessly connected resources (more details below).

*Cross-resource search: querying the databases in parallel.* In addition to the regular search, Expasy offers a ‘cross-resource search’, which allows to query a subset of web-accessible databases in parallel. Although most databases offer their own search functionality, the simultaneous query of a set of databases from a single hub is a convenient option. In this way, search results may include resources unknown to the user, yet relevant and potentially useful. This functionality, already introduced in the previous version of ExPASy (6), has been revised with the help of the respective web service providers. It now includes 19 SIB databases, such as ENZYME (21), MyHits (22), STRING(10), UniProtKB (23), ViralZone, PROSITE (24) and SWISS-MODEL Repository (see Table 2). In addition to the default behaviour (i.e., full-text search), the search engine automatically recognises certain types of formatted data, such as UniProtKB accession numbers, PDB IDs, or Ensembl IDs. As a result, it sends the query only to resources that support the specified query type. This not only

optimises search time, but also provides more relevant results.

Figure 2 shows the result page obtained with ‘COVID-19’ as input in the search box. This page consists of two parts:

- Top: the cross-resource search result (full-text mode)
- Bottom: the regular search result (resource names, keywords, categories and descriptions).

In the above example, the cross-resource search indicates that ‘COVID-19’ was found in two entries in Cellosaurus (25) and 3078 entries in UniProtKB. By clicking on the number of hits, the user is redirected to the query results in the respective databases. In case no result is returned (e.g. the format is not supported, or a resource server is temporarily down), an error icon replaces the number of results (see Figure 2, top). The regular search indicates that 19 resources relate to COVID-19, such as PROSITE, SWISS-MODEL Repository, ViralZone, Nextstrain (26), GlyConnect (27), SIB COVID-19 Integrated Knowledgebase (<https://covid-19-sparql.expasy.org/>), to name a few.

### Detailed view

Clicking on a card on the home page or on the result page shows the detailed view of the corresponding resource.

**Table 2.** Databases accessible via the cross-resource search functionality

Database name	URL	Description
SWISS-MODEL Repository SwissDock	<a href="https://swissmodel.expasy.org/repository">https://swissmodel.expasy.org/repository</a> <a href="http://www.swissdock.ch/">http://www.swissdock.ch/</a>	Protein structure homology models Docking of small ligands into protein active sites
MetaNetX Cellosaurus Bgee OMA	<a href="https://www.metanetx.org/">https://www.metanetx.org/</a> <a href="https://web.expasy.org/cellosaurus/">https://web.expasy.org/cellosaurus/</a> <a href="https://bgee.org/">https://bgee.org/</a> <a href="https://omabrowser.org/oma/home/">https://omabrowser.org/oma/home/</a>	Metabolic network repository & analysis Knowledge resource on cell lines Gene expression expertise Orthology inference among complete genomes
Selectome miROrtho OrthoDB	<a href="https://selectome.org/">https://selectome.org/</a> <a href="http://cegg.unige.ch/mirortho">http://cegg.unige.ch/mirortho</a> <a href="https://www.orthodb.org/">https://www.orthodb.org/</a>	Database of positive selection Catalogue of animal microRNA genes Ortholog evolutionary and functional annotations
neXtProt PROSITE STRING	<a href="https://www.nextprot.org/">https://www.nextprot.org/</a> <a href="https://prosite.expasy.org/">https://prosite.expasy.org/</a> <a href="https://string-db.org/">https://string-db.org/</a>	Human protein knowledgebase Protein family and domain database Protein-protein interaction networks and enrichment analysis
UniProtKB ViralZone	<a href="https://www.uniprot.org/">https://www.uniprot.org/</a> <a href="https://viralzone.expasy.org/">https://viralzone.expasy.org/</a>	Protein sequence database Fact sheets about viruses; linked to sequence databases.
ENZYME HAMAP	<a href="https://enzyme.expasy.org/">https://enzyme.expasy.org/</a> <a href="https://hamap.expasy.org/">https://hamap.expasy.org/</a>	Enzyme nomenclature database UniProtKB family classification and annotation
MyHits	<a href="https://myhits.sib.swiss/">https://myhits.sib.swiss/</a>	Relationships between protein sequences and motifs
SwissLipids VenomZone	<a href="https://www.swisslipids.org/#/">https://www.swisslipids.org/#/</a> <a href="https://venomzone.expasy.org/">https://venomzone.expasy.org/</a>	Knowledge resource for lipids Portal to venom protein UniProtKB entries

1

Cross-reference search results

Display of the results if the query is meaningless with respect to a specific database

Number of hits found in the subset of SIB databases

2

Regular search results

**ExPASy** Home About SIB News Contact

COVID-19  
e.g. BLAST, UniProt, MSH6, Albumin...

**Hits found in 19 SIB databases for "COVID-19"**

Structural Biology	Genes & Genomes	Proteins & Proteomes
<b>SWISS-MODEL Repository</b> Protein structure homology models 6 hits	<b>Bgee</b> Gene expression expertise 0 hit	<b>neXtProt</b> Human protein knowledgebase 6 hits
<b>SwissDock</b> Docking of small ligands into protein active sites 0 hit	<b>Evolution &amp; Phylogeny</b>	<b>PROSITE</b> Protein family and domain database 1 hit
<b>System Biology</b>	<b>OMA</b> Orthology inference among complete genomes 0 hit	<b>STRING</b> Protein-protein interaction networks and enrichment analysis 0 hit
<b>MetaNetX</b> Metabolic network repository & analysis 0 hit	<b>OrthoDB</b> Ortholog evolutionary and functional annotations 0 hit	<b>UniProtKB</b> Protein-sequence database 3354 hits
<b>Cellosaurus</b> Knowledge resource on cell lines 2 hits	<b>miROrtho</b> Catalogue of animal microRNA genes 0 hit	<b>ViralZone</b> Fact sheets about viruses; linked to sequence databases. 0 hit
<b>Selectome</b> Database of positive selection 0 hit	<b>ENZYME</b> Enzyme nomenclature database 0 hit	<b>SwissLipids</b> Knowledge resource for lipids 0 hit
		<b>HAMAP</b> UniProtKB family classification and annotation 0 hit
		<b>MyHits</b> Relationships between protein sequences and motifs 0 hit
		<b>VenomZone</b> Portal to venom protein UniProtKB entries 0 hit

**19 resources found for "COVID-19"**

<b>PROSITE</b> Protein family and domain database	<b>SWISS-MODEL Repository</b> Protein structure homology models	<b>HAMAP</b> UniProtKB family classification and annotation	<b>ViralZone</b> Fact sheets about viruses; linked to sequence databases.
<b>neXtProt</b> Human protein knowledgebase	<b>Cellosaurus</b> Knowledge resource on cell lines	<b>Nextstrain</b> Impact of pathogen genome data on science and public health	<b>GlyConnect</b> Integrated data platform to study glycosylation

**Figure 2.** ExPASy result page: the two types of search results are visually separated. 1. (Top) Cross-resource search result; with the number of hits found in each of the 19 databases; 2. (Bottom) Regular search.



The screenshot shows the ExPASy website interface for the ViralZone resource. At the top, there is a search bar with a magnifying glass icon and a red search button. Below the search bar, there are navigation links for Home, About, SIB News, and Contact. The main heading is 'ViralZone'. Below the heading, there is a description of the resource and a video link. Two grey boxes highlight EDAM terms: 'What you can do with this resource?' (Operation EDAM terms) and 'Browse these keywords in ExPASy' (Topic, Data, and Format EDAM terms). A section titled 'You might also be interested in' lists several related resources like OrthoDB, ImmunoDB, UniProtKB/Swiss-Prot, UniProtKB, HAMAP-Scan, HAMAP, SwissOrthology, and TASmania. Dashed lines connect the EDAM terms to the related resources.

**Figure 3.** Detailed view of a resource (ViralZone). The two grey boxes display the EDAM (28) terms related to *Operations* [1] and *Topics, Data and Formats* [2]. The bottom of the page [3] lists the suggestions of resources that share at least one term of the *Topic* type with ViralZone.

Figure 3 shows a detailed view of ViralZone, a web resource for all virus genera and families. Overall, the following information describes each resource:

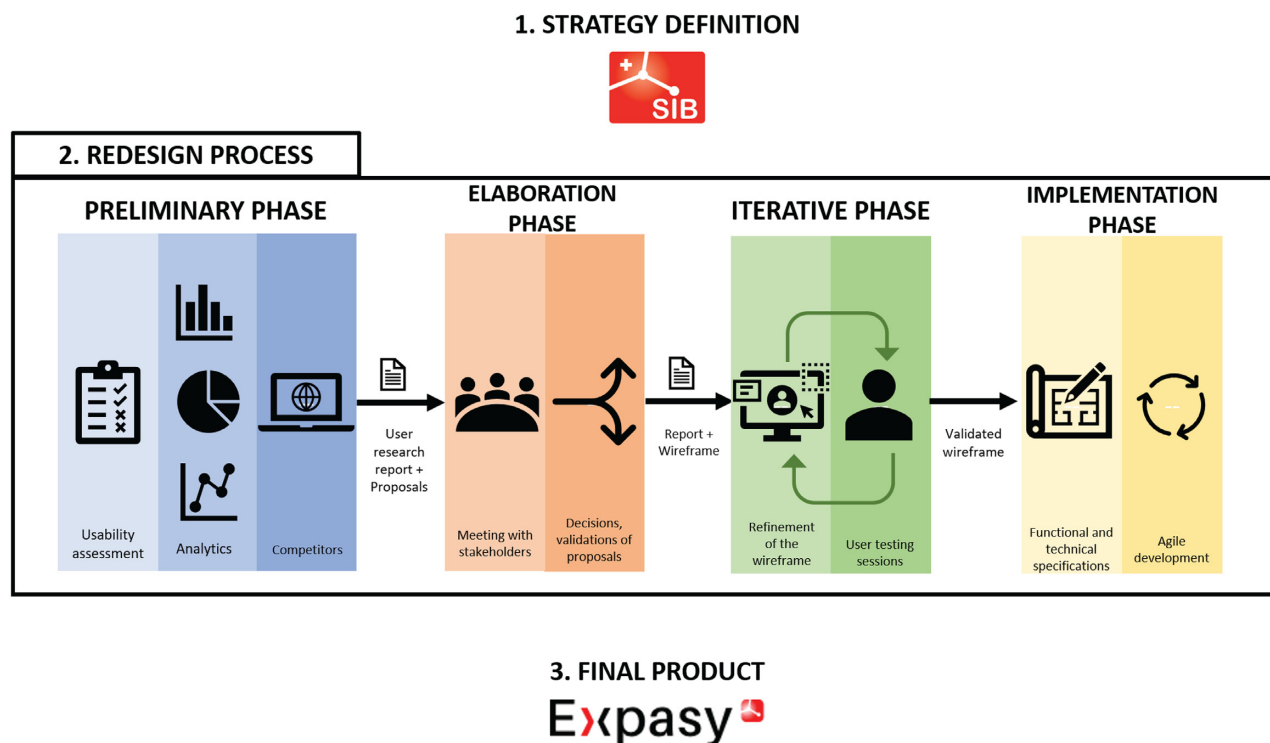
- Name
- Description
- Type of resource (software tool and/or database)
- Categories (Genes & Genomes, Genomics, Metagenomics, Transcriptomics, Proteins & Proteomes, Evolution & Phylogeny, Evolution biology, Population genetics, Structural Biology, Drug design, Medicinal chemistry, Structural analysis, Systems Biology, Glycomics, Lipidomics, Metabolomics or Text mining & Machine learning)
- Optional presentation video
- URL
- SIB group operating the resource
- Licence terms
- Keywords of type:
  - 'What you can do with this resource?'
  - 'Browse these keywords in ExPASy'
- Resources of interest for the user shown as cards

Clicking on a keyword in the right box triggers a new query matching that keyword. Keywords in the old ExPASy were not standardised or following any commonly used ontology (those keywords will be designated as 'in-house keywords' in the rest of the article). In the new ver-

sion, keywords comply with EDAM (28), a comprehensive ontology of well-established concepts in bioinformatics and computational biology. EDAM applicability to searching, categorising and automatic handling of resources has been validated by implementations in eSysbio (<https://nsls.bioinfo.no/>), Bio-jETI (29) and EMBOSS (30), demonstrating its relevance to resource catalogues. The ontology is actively developed, maintained and supported by a team of specialists. The four sub ontologies of EDAM are used: Operation, Topic, Data and Format. In EDAM, 'Operation' is 'A function that processes a set of inputs and results in a set of outputs'. Thus, all EDAM terms of the 'Operation' type are displayed in the box 'What you can do with this resource?'. By clicking on one of the EDAM terms in this box, the user can retrieve all resources that perform the same task. 'Topic', 'Data' and 'Format' define, respectively, the domain of application, the type of information and the data format used or output by the resource. Those terms are shown in the box 'Browse these keywords in ExPASy'. The list of resources that share at least one EDAM term of the 'Topic' type is indicated in the 'You might also be interested in' section, which allows the user to explore SIB resources (see Figure 3, bottom, for details).

## THE REDESIGN PROCESS: MOTIVATION, STRATEGY AND DEVELOPMENT

The redesign process started in 2019. Its aim was to improve user-friendliness, visual identity and content, as well as its



**Figure 4.** Process of development of Expasy. 1. Project strategy definition, 2. Four-phased redesign process (preliminary phase, elaboration phase, iterative phase, implementation phase) and 3. Final product.

responsiveness on mobile devices. Additionally, we seized the opportunity to change the casing of the word ExPASy to Expasy to distinguish it from the original purpose of the website (the study of proteins) while keeping a world-renowned brand in bioinformatics. We followed a UCD approach for the redesign, anticipating that optimised user interfaces and efficient organisation of information are more conducive to scientific discoveries. Yet, professionals usually tend to wrongly assume that end-users share the same view and behave in a similar way in a given situation. In reality, the development of high scientific quality resources used to their full potential and performing adequately, must include end-users in the design process from the very beginning. With this in mind, the redesign of Expasy was carried out through four developmental phases that are described below (see Figure 4 for an illustration of the process).

### Preliminary phase

In User Experience (UX), the preliminary phase is essential for making informed decisions. To this end, we first assessed the old ExPASy according to usability, usage statistics of the website as well as user behaviour. We also evaluated a series of well-established portals in the world of life sciences, the so-called ‘competitors’.

This preliminary phase allowed us to gain significant insight into best practices and user experience improvement according to the following four criteria: usability, usage, user behaviour and benchmarking. The outcome of our evaluation is summarised in Table 3.

*Usability.* Overall, we identified two major usability issues in the old ExPASy website (Figure 5). Firstly, users did not distinguish search types (Figure 5A). Prior to launching a search, the user was expected to select the type of search (regular or cross-resource) from a drop-down menu. Our analyses and user tests showed that very few users were aware of this option. As a consequence, users typically ran the default search (cross-resource search in 77% of cases, according to the server logs). Secondly, the website was not responsive, that is, not optimized for mobile devices (Figure 5B). Not only did this lead to very few users accessing the platform through mobile devices (around 6%), it also penalised indexing by the major search engines.

*Usage.* The study of the old ExPASy usage figures were based on Google Analytics spanning the year 2018. Overall, we observed a bounce rate of 50% which is considered as good. A bounce is when the user opens a single page on a website and exits without triggering any other action. There was an almost exclusive use of the website through desktops (94%), a huge number of acquisitions by direct access (32%), an average number of acquisitions by organic (search engine) searches (56%), and a slightly lower acquisition by referral from other, external websites (21%). Usage figures are comprehensively provided in Supplementary File S1.

*User behaviour.* A close look at the user behaviour flows revealed the low level of navigation between resources. From this observation we identified the need for further development to support navigation and exploration. Supplementary File S1 also highlights this trend.

**Table 3.** The main outcomes and insights of the preliminary phase

Type of study	Issue	Solution/insight
<b>Usability</b>	Selection of search type (cross-resource or regular) from a dropdown menu.	Provide a single search bar for all types of search, with all types of result on a joint page.
	Website not supported by mobile devices	Make the website responsive.
<b>Usage</b>	Very limited access by mobile devices	Make the website responsive.
	Average number of acquisitions by organic searches	Improve the indexing by search engines by (i) making the website responsive; (ii) ensuring that no URL is left with a 404 (page not found) response code; and (iii) adding more content to Expasy, using controlled vocabularies (helps search engines to understand the expertise areas of Expasy).
<b>User behaviour</b>	Low number of acquisitions by referrals	Increase the number of backlinks to Expasy from various partner websites.
	Lack of navigation between resources	Ease navigation and exploration by creating a network of functionally- related resources.
<b>Competitor benchmarking</b>	EMBL-EBI and NCBI	-
	EMBL-EBI and NCBI	-
	NCBI	-
	bio.tools	-
	BioCatalogue	-
		Provide a single search bar (for cross-resource and regular searches). This is an example of an assumption we could validate.
		Provide a joint results page (cross-resource and regular search results). This illustrates how we took advantage of the familiarity factor.
		Provide filters to narrow down the number of resources. This is an example of an assumption we could validate.
		Use an established ontology to describe the resources. This is an example of adopting best-practices.
		Involve the resource providers in the description of resources and the continuous revision of information. This contributed to identifying the added-value of Expasy.

*Competitor benchmarking.* Benchmarking is an efficient way to analyse user interfaces and community standards. In this work, we targeted well established websites to (i) evaluate ongoing practices in the field of bioinformatics, (ii) benefit from the familiarity factor, (iii) validate or disprove assumptions and (iv) assess the added value of Expasy compared to other bioinformatics portals. For this study, we focused on four portals:

1. The EMBL-EBI website (<https://www.ebi.ac.uk/>)
2. The NCBI website (<https://www.ncbi.nlm.nih.gov/>)
3. The bio.tools portal (<https://bio.tools/>)
4. The Biocatalogue (<https://www.biocatalogue.org/>)

The details of this benchmarking can be found in Supplementary File S2. Examples can be found in Table 3.

The analysis of the old ExpASy usage numbers as well as the user behaviour flows led to the creation of personas, that is, fictional characters, which are stand-ins for the different Expasy user types with characteristics, needs and expectations. Creating personas helps software developers to (i) assess the extent of different needs and expectations, and (ii) optimize the user experience (31). Three personas were created: a bioinformatician, a junior, and a senior biologist (Supplementary File S3 represents the personas and illustrates to what extent the outcome of the study supports the features of each persona).

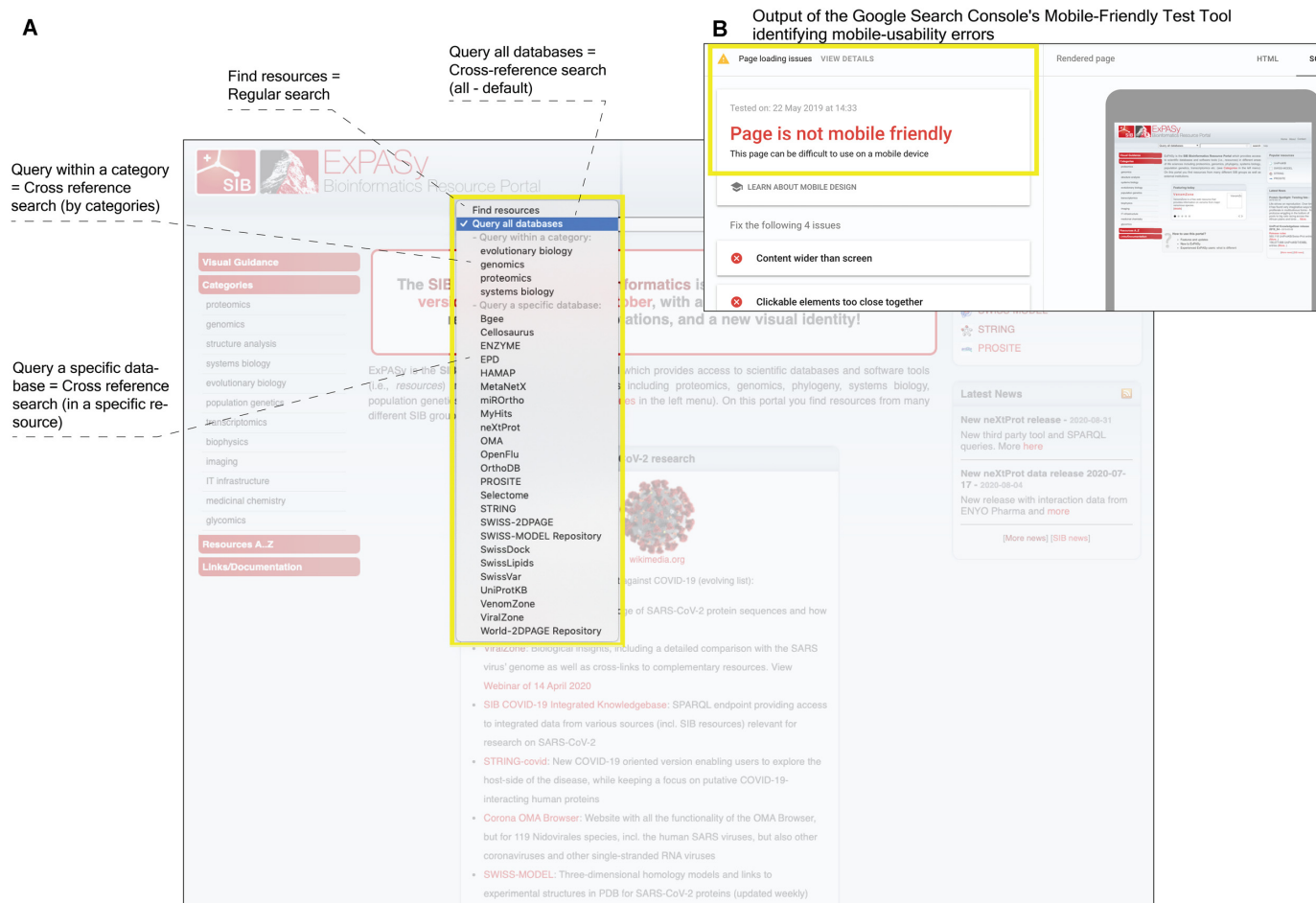
Finally, the new Expasy was drafted as a wireframe as shown in Figure 6. A wireframe is a visual guide that represents the skeleton of a website. It depicts the layout of the user interface elements and their interactions. It lacks ty-

pographical style, colour or graphics, as the main focus is functionality, behaviour and content. In other words, it focuses on what a screen does, not on its appearance. A wireframe is employed in user tests, allowing for early feedback on products. Usually, a wireframe is iteratively enhanced at each user test to produce an optimised prototype used by developers for implementation.

The first wireframe consisted of (i) a single search bar, performing the two types of search in parallel (regular and cross-resource search) with no prior selection from a drop-down list, (ii) a homepage showing all resources listed as small cards, with a short description and additional information (website, contact, type, etc.), (iii) different types of filters (categories, resource type, etc.) to explore SIB resources according to various criteria, (iv) a detailed view of each resource providing the user with more information, and a 'You might also be interested in...' section allowing to broaden the scope to other SIB resources.

At this stage, the pending issues were:

- What information should the cards contain?
- What information should the detailed views contain?
- How to ensure links between resources with common features?
- Which filters should be applied?
- Which resource categories should be used to cover the diversity of SIB resources?
- How to monitor revision and long-term update of resource description?



**Figure 5.** Usability issues of the previous version of ExPASy. (A) Dropdown menu (regular search and cross-resource search) and (B) output of the Google Search Console's mobile-friendly test tool.

### Elaboration phase

For the elaboration phase, we organised an interactive workshop, which brought together 17 members of the SIB community, including PIs of major SIB resources, software developers, scientists, communication and training specialists. Indeed, the SIB community brings together both the developers or providers of the resources, and the ExPASy end-users. Furthermore, the workshop aimed to ease the adoption of the future version of ExPASy, which is crucial for the long-term revision of resource descriptions, and more generally to sustainable content quality.

Various exercises were carried out during the workshop, and their outcome was processed in order to refine the wireframe.

More specifically, the audience unanimously confirmed the following prerequisites:

- A unique search box performing the two search types (regular and cross-resource search),
- The exclusion of citations, contact and website icon in the resource detailed view, and a clear description of the resource purpose,
- The replacement of in-house keywords by EDAM terms,
- The only indispensable filter is by category.

The field of bioinformatics evolves rapidly, therefore, through a 'card-sorting' exercise, the categories and sub-categories were updated, as shown in Table 4. At the end of the workshop, a new version of the wireframe was released and validated by the participants. This wireframe, resulting from a consensus of the SIB community, was then proposed to a set of external users.

### Iterative phase

The iterative phase is about meeting the users, showing them the wireframe and improving it based on their feedback. Five users were recruited either through a mailing campaign across the authors of scientific articles citing ExPASy or in person at various scientific meetings. It was established that testing a wireframe on no more than five users is sufficient to identify the major usability issues (65–85%) (32). In practice, the redundancy of user feedback is such that all important issues are brought out with the first five users and very few new ones arise when more users are consulted. The 'law of diminishing returns' (33) applies. Conversely, testing a wireframe on a single user is risky: a single person may perform actions by accident or in an unrepresentative manner and this may skew the test results. During the tests,



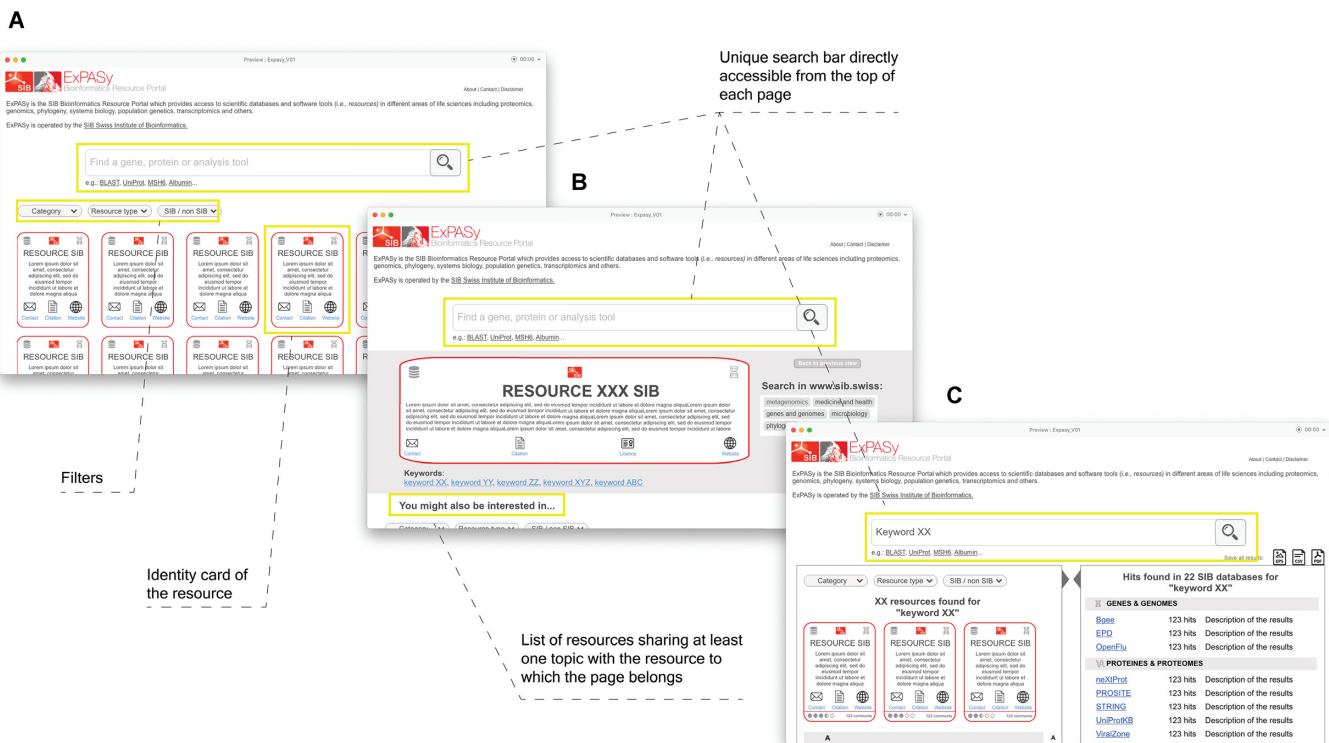


Figure 6. The first wireframe of the ExPASy project. (A) Home page, (B) resource detailed view and (C) search result page.

Table 4. Categories and sub-categories in ExPASy

Category	Sub-category
Genes & Genomes	Genomics
	Metagenomics
	Transcriptomics
Proteins & Proteomes	
Evolution & Phylogeny	Evolution biology
	Population genetics
Structural Biology	Drug design
	Medicinal chemistry
Systems Biology	Structural analysis
	Glycomics
	Lipidomics
Text mining & Machine learning	Metabolomics

users were asked to navigate the old ExPASy as well as different versions of the wireframe. Simultaneously, they were invited to compare and comment on the diverse user experiences. To avoid the anchoring bias (34), the different versions of the website were submitted in a random order. Here are some examples of exercises proposed to the users:

- Find a database collecting expert curated knowledge about glycan bindings of human pathogens
- Retrieve all the resources belonging to the ‘Evolutionary Biology’ category

- Retrieve information on the MSH6 gene. In which tissue is this gene expressed in human?

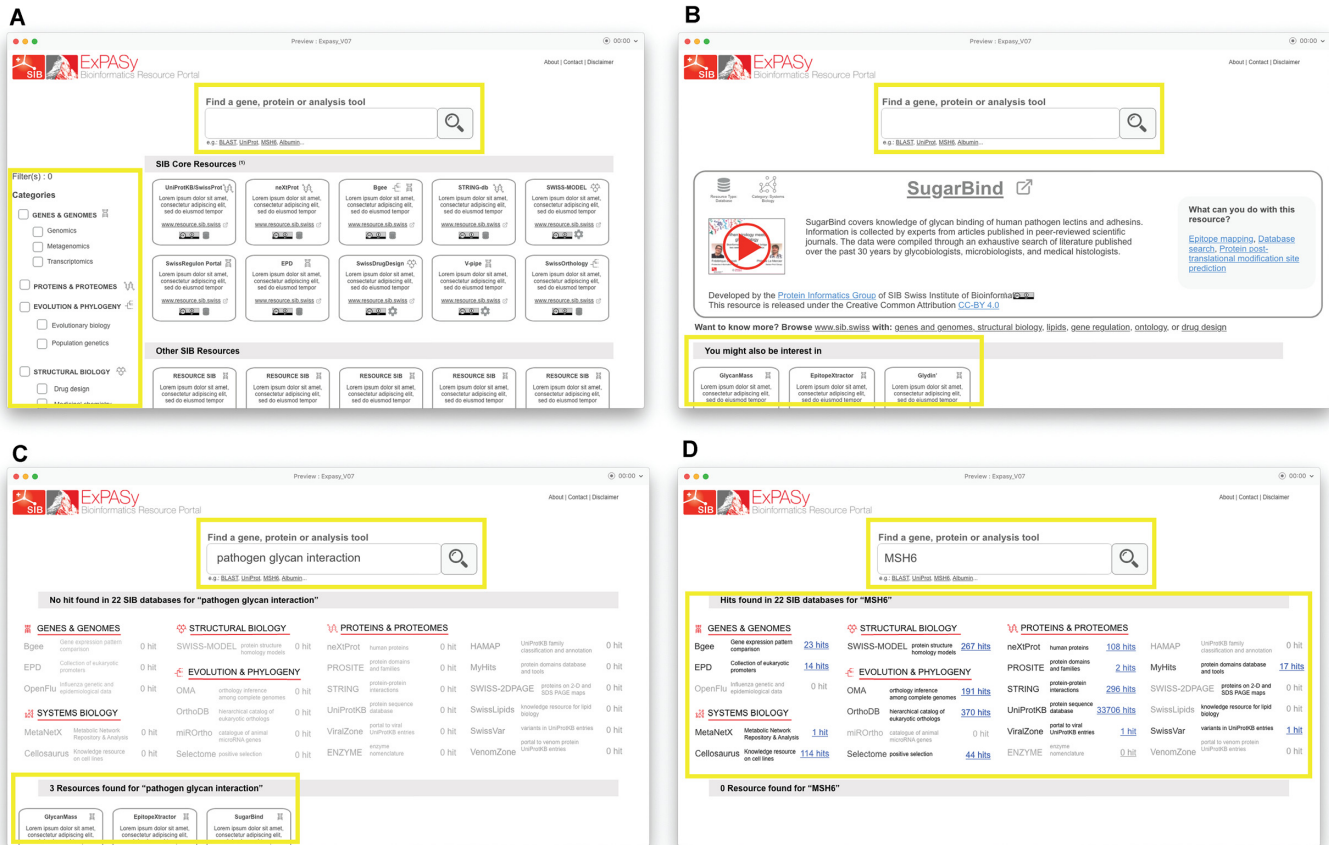
At the end of the iterative phase, the outcome that led to the final wireframe version (Figure 7) was the following:

- Overall, the users liked the unified search box and confirmed their unawareness of the regular search in the old implementation.
- The added value of ExPASy lies in the classification of the resources by category. Besides, users preferred the panel on the left, over the initially proposed dropdown menu at the top.
- The result page, with its two types of results (Figure 7C and D), was well understood. More specifically, in the cross-resource search results, the users appreciated the classification of results by category and the short description of each queried database.

### Implementation phase

The implementation phase spans stages from the revision of the pre-existing ExPASy content to the development of the new product before going live.

To update the pre-existing data, we started by removing decommissioned resources as well as non-SIB resources. In the old ExPASy, in-house keywords were used to describe each resource main features. Manual mapping of those keywords to EDAM terms was performed and applied to each resource. The mapping result was then sent to the resource providers along with additional information such as the



**Figure 7.** The final version of the iteratively refined wireframe as validated by the end-users. (A) Home page; (B) resource detailed view; (C and D) search result pages: C highlights on the regular search, D highlights on the cross-resource search. Note that the search bar is always present.

long and short descriptions, the contact email or the URL. Each of the resource descriptions was reviewed and enriched by the resource providers. A total of 154 resources were re-evaluated before the new Expasy implementation went live.

As for the development of the new release, a few key points had to be considered:

- Knowing that about 49% of global internet traffic can be attributed to mobile devices (source: <https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/worldwide>, October 2020), Expasy needed to be responsive, not only to attract more mobile device users, but also to improve indexing by the major search engines,
- Work on the Search Engine Optimization to increase the number of newcomers via organic searches,
- Update the visual identity (new stylesheet, new logo),
- Ensure the redirection of about 5000 URLs to keep existing users,
- Provide an administration interface to add/update/delete resources, open to resource providers and administrators only.

To increase speed of development and ensure user satisfaction, we opted for an agile methodology (<https://agilemanifesto.org>). In software development, agile refers

to discovering requirements and developing solutions through a high level of engagement of cross-functional teams, enabling early delivery and continuous improvement. Eleven people (developers, business analysts, product owner, SEO expert, user interface designer, graphic designer, system administrators) participated in the implementation phase, and the new Expasy went live on 15 October 2020. The resource providers were granted access to the Expasy administration interface and provided with a video tutorial. As of January 2021, 15 new resources have been added, and 50 have been updated. Involving resource providers in the continuous content update of Expasy is crucial for ensuring accurate and relevant information. With this procedure, along with the yearly update reminder sent to resource providers, we have optimised access to up-to-date and reliable data.

## CONCLUSION

The new version of Expasy was released in October 2020. Since then, an increase of 15% in the number of daily users, compared to the same period last year, is observed. The traffic from mobile devices has increased from 6 to 10%. In the future, the SIB community will strive to provide more resources to Expasy users while ensuring that pre-existing information remains up to date. The new version of Expasy was intended as an evolving web application with a grow-

ing number of high-quality interconnected resources that forms a solid ground on which the life science community can stand. The portal is a gateway to bioinformatics, available to expert users, beginner researchers, teachers, students and more. The new version of Expasy aims to create a satisfactory user experience. Our user-centric approach was successfully applied in all phases of development and will remain our strategy in future implementations.

## DATA AVAILABILITY

Expasy is available at [www.expasy.org](http://www.expasy.org). This website is free and open to all users and there is no login requirement.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We are particularly grateful to the SIB-cofounders Ron Appel and Amos Bairoch who created the first version of EXPASy almost three decades ago. Expasy would not exist without the resource providers: we take this opportunity to thank them for their immense effort in developing, maintaining and revising the resources. We also want to thank all the people who contributed to the design, implementation and operation of this new version of Expasy. In particular, we thank the participants of the workshop in 2019 and the internal groups at SIB (Core-IT, LTTO and Communications and Scientific Events departments). Lastly, we would like to warmly thank the users who participated in the user testing sessions and those who have been faithful for decades.

## FUNDING

SIB Swiss Institute of Bioinformatics gratefully acknowledges the State Secretariat for Education, Research and Innovation SERI. Funding for open access charge: SIB Swiss Institute of Bioinformatics.

*Conflict of interest statement.* None declared.

## REFERENCES

- Agarwala,R., Barrett,T., Beck,J., Benson,D.A., Bollin,C., Bolton,E., Bourexis,D., Brister,J.R., Bryant,S.H., Canese,K. *et al.* (2018) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **46**, D8–D13.
- Cook,C.E., Lopez,R., Stroe,O., Cochrane,G., Brooksbank,C., Birney,E. and Apweiler,R. (2019) The European Bioinformatics Institute in 2018: Tools, infrastructure and training. *Nucleic Acids Res.*, **47**, D15–D22.
- Ison,J., Ienasescu,H., Chmura,P., Rydza,E., Menager,H., Kalaš,M., Schwammle,V., Gruning,B., Beard,N., Lopez,R. *et al.* (2019) The bio.tools registry of software tools and data resources for the life sciences. *Genome Biol.*, **20**, 164.
- Bhagat,J., Tanoh,F., Nzuobontane,E., Laurent,T., Orlowski,J., Roos,M., Wolstencroft,K., Alekseyevs,S., Stevens,R., Pettifer,S. *et al.* (2010) BioCatalogue: A universal catalogue of web services for the life sciences. *Nucleic Acids Res.*, **38**, W689–W694.
- Appel,R.D., Bairoch,A. and Hochstrasser,D.F. (1994) A new generation of information retrieval tools for biologists: the example of the EXPASy WWW server. *Trends Biochem. Sci.*, **19**, 258–260.
- Artimo,P., Jonnalagedda,M., Arnold,K., Baratin,D., Csardi,G., De Castro,E., Duvaud,S., Flegel,V., Fortier,A., Gasteiger,E. *et al.* (2012) ExpASy: SIB bioinformatics resource portal. *Nucleic Acids Res.*, **40**, W597–W603.
- SIB Swiss Institute of Bioinformatics Members (2016) The SIB Swiss Institute of bioinformatics' resources: focus on curated databases. *Nucleic Acids Res.*, **44**, D27–D37.
- Poux,S., Arighi,C.N., Magrane,M., Bateman,A., Wei,C.-H., Lu,Z., Boutet,E., Bye-A-Jee,H., Famiglietti,M.L., Roehert,B. *et al.* (2017) On expert curation and scalability: UniProtKB/Swiss-Prot as a case study. *Bioinformatics*, **33**, 3454–3460.
- Zahn-Zabal,M., Michel,P.A., Gateau,A., Nikitin,F., Schaeffer,M., Audot,E., Gaudet,P., Duek,P.D., Teixeira,D., De Laval,V.R. *et al.* (2020) The neXtProt knowledgebase in 2020: Data, tools and usability improvements. *Nucleic Acids Res.*, **48**, D328–D334.
- Szklarczyk,D., Gable,A.L., Lyon,D., Junge,A., Wyder,S., Huerta-Cepas,J., Simonovic,M., Doncheva,N.T., Morris,J.H., Bork,P. *et al.* (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, **47**, D607–D613.
- Bastian,F.B., Roux,J., Niknejad,A., Comte,A., Fonseca Costa,S.S., de Farias,T.M., Moretti,S., Parmentier,G., de Laval,V.R., Rosikiewicz,M. *et al.* (2021) The Bgee suite: integrated curated expression atlas and comparative transcriptomics in animals. *Nucleic Acids Res.*, **49**, D831–D847.
- Bienert,S., Waterhouse,A., De Beer,T.A.P., Tauriello,G., Studer,G., Bordoli,L. and Schwede,T. (2017) The SWISS-MODEL repository-new features and functionality. *Nucleic Acids Res.*, **45**, D313–D319.
- Daina,A. and Zoete,V. (2019) Application of the SwissDrugDesign online resources in virtual screening. *Int. J. Mol. Sci.*, **20**, 4612.
- Pavelin,K., Cham,J.A., de Matos,P., Brooksbank,C., Cameron,G. and Steinbeck,C. (2012) Bioinformatics meets user-centred design: a perspective. *PLoS Comput. Biol.*, **8**, e1002554.
- Karamanis,N., Pignatelli,M., Carvalho-Silva,D., Rowland,F., Cham,J.A. and Dunham,I. (2018) Designing an intuitive web application for drug discovery scientists. *Drug Discov. Today*, **23**, 1169–1174.
- Posada-Céspedes,S., Seifert,D., Topolsky,I., Jablonski,K.P., Metzner,K. and Beerenwinkel,N. (2021) V-pipe: a computational pipeline for assessing viral genetic diversity from high-throughput data. *Bioinformatics*, <https://doi.org/10.1093/bioinformatics/btab015>.
- Hulo,C., De Castro,E., Masson,P., Bougueleret,L., Bairoch,A., Xenarios,I. and Le Mercier,P. (2011) ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.*, **39**, D576–D582.
- Liechti,R., Gleizes,A., Kuznetsov,D., Bougueleret,L., Le Mercier,P., Bairoch,A. and Xenarios,I. (2010) OpenFluDB, a database for human and animal influenza virus. *Database (Oxford)*, **2010**, baq004.
- Noll,N.B., Aksamentov,I., Druelle,V., Badenhurst,A., Ronzani,B., Jefferies,G., Albert,J. and Neher,R.A. (2020) COVID-19 scenarios: an interactive tool to explore the spread and associated morbidity and mortality of SARS-CoV-2. medRxiv doi: <https://doi.org/10.1101/2020.05.05.20091363>, 12 May 2020, preprint: not peer reviewed.
- Gormley,C. and Tong,Z. (2015) Elasticsearch: the definitive guide. <https://dl.acm.org/doi/10.5555/2904394>.
- Bairoch,A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.
- Pagni,M., Ioannidis,V., Cerutti,L., Zahn-Zabal,M., Jongeneel,C.V., Hau,J., Martin,O., Kuznetsov,D. and Falquet,L. (2007) MyHits: improvements to an interactive resource for analyzing protein sequences. *Nucleic Acids Res.*, **35**, 433–437.
- The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
- Sigrist,C.J.A., De Castro,E., Cerutti,L., Cuče,B.A., Hulo,N., Bridge,A., Bougueleret,L. and Xenarios,I. (2013) New and continuing developments at PROSITE. *Nucleic Acids Res.*, **41**, D344–D347.
- Bairoch,A. (2018) The cellosaurus, a cell-line knowledge resource. *J. Biomol. Tech.*, **29**, 25–38.
- Hadfield,J., Megill,C., Bell,S.M., Huddleston,J., Potter,B., Callender,C., Sagulenko,P., Bedford,T. and Neher,R.A. (2018)

- Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, **34**, 4121–4123.
27. Alocci,D., Mariethoz,J., Gastaldello,A., Gasteiger,E., Karlsson,N.G., Kolarich,D., Packer,N.H. and Lisacek,F. (2019) GlyConnect: glycoproteomics goes visual, interactive, and analytical. *J. Proteome Res.*, **18**, 664–677.
  28. Ison,J., Kalaš,M., Jonassen,I., Bolser,D., Uludag,M., McWilliam,H., Malone,J., Lopez,R., Pettifer,S. and Rice,P. (2013) EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. *Bioinformatics*, **29**, 1325–1332.
  29. Lamprecht,A.L., Margaria,T. and Steffen,B. (2009) Bio-jETI: a framework for semantics-based service composition. *BMC Bioinformatics*, **10**, S8.
  30. Rice,P., Longden,L. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
  31. Gabella,C. and Duvaud,S. (2020) Applying User Experience (UX) methodologies to science: the example of the BioSODA user interface. Zenodo doi: <https://doi.org/10.5281/ZENODO.4155503>, 29 October 2020, preprint: not peer reviewed.
  32. Nielsen,J. and Landauer,T.K. (1993) A Mathematical model of the finding of usability problems. In: *Conference on Human Factors in Computing Systems - Proceedings*. ACM, NY, pp. 206–213.
  33. Shephard,R.W. and Färe,R. (1974) The law of diminishing returns. *Zeitsch. Natl. J. Econ.*, **34**, 69–90.
  34. Furnham,A. and Boo,H.C. (2011) A literature review of the anchoring effect. *J. Socio. Econ.*, **40**, 35–42.