BMC Structural Biology

DATABASE Open Access

REFOLDdb: a new and sustainable gateway to experimental protocols for protein refolding

Hisashi Mizutani[1], Hideaki Sugawara[1*], Ashley M. Buckle[2], Takeshi Sangawa[3], Ken-ichi Miyazono[4], Jun Ohtsuka[4], Koji Nagata[4], Tomoki Shojima[4], Shohei Nosaki[4], Yuqun Xu[4], Delong Wang[4], Xiao Hu[4], Masaru Tanokura[4] and Kei Yura[1,5,6,7]

Abstract

Background: More than 7000 papers related to "protein refolding" have been published to date, with approximately 300 reports each year during the last decade. Whilst some of these papers provide experimental protocols for protein refolding, a survey in the structural life science communities showed a necessity for a comprehensive database for refolding techniques. We therefore have developed a new resource – "REFOLDdb" that collects refolding techniques into a single, searchable repository to help researchers develop refolding protocols for proteins of interest.

Results: We based our resource on the existing REFOLD database, which has not been updated since 2009. We redesigned the data format to be more concise, allowing consistent representations among data entries compared with the original REFOLD database. The remodeled data architecture enhances the search efficiency and improves the sustainability of the database. After an exhaustive literature search we added experimental refolding protocols from reports published 2009 to early 2017. In addition to this new data, we fully converted and integrated existing REFOLD data into our new resource. REFOLDdb contains 1877 entries as of March 17th, 2017, and is freely available at http://p4d-info.nig.ac.jp/refolddb/.

Conclusion: REFOLDdb is a unique database for the life sciences research community, providing annotated information for designing new refolding protocols and customizing existing methodologies. We envisage that this resource will find wide utility across broad disciplines that rely on the production of pure, active, recombinant proteins. Furthermore, the database also provides a useful overview of the recent trends and statistics in refolding technology development.

Keywords: Solubilization, Inclusion body, Refolding, Renaturation, Crystallization

Background

Establishment of heterologous expression technology of recombinant proteins has revolutionized protein purification such that it is performed with cloned, recombinant proteins expressed in a suitable host. The predominant host is *Escherichia coli*. However, many overexpressed proteins in *E. coli* are found in an insoluble form called inclusion bodies (IBs). Since the target protein is often highly pure in washed IBs, the challenge is not so much to purify the target, but rather to solubilize IBs and refold the protein into its native, biologically active state [1, 2]. While many of the operations to prepare IBs are quite general—expression, cell disruption, IB isolation and washing, the precise conditions that are required to achieve efficient refolding vary for each protein.

The refolding experiments consist of two steps: (1) the solubilization of IBs by adding a denaturant and (2) the renaturation of the denatured protein by lowering the denaturant concentration. The solubilization step is relatively easily, performed by adding a denaturant, typically urea or guanidinium chloride at a final concentration of 6–8 M or 6 M, respectively. The renaturation step is often difficult. In order to maximize the refolding yield,

* Correspondence: hsugawar@nig.ac.jp
[1]Center for Information Biology, National Institute of Genetics, 1111 Yata Mishima, Shizuoka 411-8540, Japan
Full list of author information is available at the end of the article

the optimization of the following experimental methods/ conditions of this process is required:

(a) refolding method: dilution and dialysis [3], gel filtration column chromatography, column adsorption and desorption [4], and high pressure [5] represent the most common methods. These methods are used to lower the denaturant concentration and allow protein refolding in an aqueous buffer.
(b) pH: In general, pI should not be used for refolding experiment to avoid isoelectric point precipitation [6].
(c) temperature: Temperature has an effect on the stability and mobility of the refolding intermediates [7].
(d) protein concentration: Protein concentration determines the degree of "crowding" and thus the frequency of molecular collisions between the unfolded molecules as well as folding intermediates, which can promote aggregation [8].
(e) additive(s): Some compounds may stabilize the refolding intermediates and avoid aggregation [9].

Because the suitable refolding methods/conditions differ from protein to protein, a knowledge database of optimized refolding methods/conditions for each protein is an important resource for many biochemists and molecular biologists. Thus, the REFOLD database established and published by Monash University in 2006 played an important role as the sole information source for refolding experiments [10–13]. This database, however, suspended its updates in 2009. We carried out a preliminary study in 2013 for the development of a sustainable database on protein refolding technologies. We decided that a new database, REFOLDdb, was required as a gateway to experimental methodologies that describe experimental refolding in detail. We therefore designed a simple data format and consistent data representation among entries so that users are able to easily interrogate the database and painlessly retrieve and understand search results. The design also allows straightforward maintenance, allowing the database to be sustainable over a long period.

The sustainability of biological databases is a serious issue [14, 15] and database developers have to analyze cost-effectiveness in advance. In the case of databases relating to technologies (Tech_db), the data volume will not expand as rapidly as in the case of molecular databases, for example the International Sequence Database [16], the Worldwide Protein Data Bank [17], UniProt [18], SUPERFAMILY database [19]. Nevertheless, developer of Tech_db must be sensitive to the direct and indirect cost of data extraction from the primary articles, curation and updating. REFOLDdb is designed to balance both cost and usefulness.

We have captured the refolding data from up-to-date literature as well as retrospectively from articles published since 2009. We also updated, converted and integrated the data stored in the REFOLD database into REFOLDdb. As of March 17th, 2017, REFOLDdb provides users with data on 1877 experimental methods for refolding 1628 proteins. Most of these data were extracted from 1232 publications.

Construction and content

We searched the NCBI PubMed database by a keyword search of "(refolding[All Fields] OR renaturation[All Fields]) AND ("proteins"[MeSH Terms] OR "proteins"[All Fields] OR "protein"[All Fields])" to find 2606 research reports published between 2009-early 2017 that might be relevant to REFOLDdb. Manual inspection of the results identified 420 reports that contained experimental protocols for the refolding of 650 proteins. These data were then integrated in REFOLDdb along with the data stored in the REFOLD database. REFOLDdb refers to 1232 publications in total (Full list available via a menu "List of publications referred by REFOLDdb" in "About" page at http://p4d-info.nig.ac.jp/refolddb/about.cgi?lang=EN).

Due to the standardization and other extension of the data format, the database now contains the following functionality: (1) it is searchable by sequence similarity; (2) it is equipped with statistics that enables the discovery of trends in refolding techniques; and (3) it is easy to upload/submit new data to the database manager. Specifically, the database has the following three sections: Article [title/abstract/PubMed ID/Author/Journal/Date], Protein [Protein name/Amino acid sequence/Comment/UniProt ID/Function/Domain] and Experiment [Refolding methods/pH/Temperature/Validation]. We did not itemize "protein concentration" and "additive(s)", because "protein concentration" is often missing in articles and the description of "additive(s)" is quite heterogeneous. REFOLDdb is composed of 12 tables in a relational database system.

REFOLDdb was created using open-source PostgreSQL relational database server software version 9.2.14 (https://www.postgresql.org/), running under CentOS 7 Server (version 7.2-1511) on a virtual machine based on VMware ESXi (http://www.vmware.com/products/esxi-and-esx.html). The system complies with the security policy of the National Institute of Genetics, Japan. A web-based query interface to the database was developed using the Perl programming language and PDO database abstraction classes (http://jp2.php.net/manual/en/book.pdo.php), and is hosted on the same virtual machine running the Apache 2.4.6 web server.

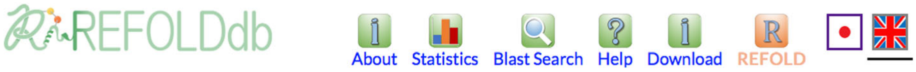


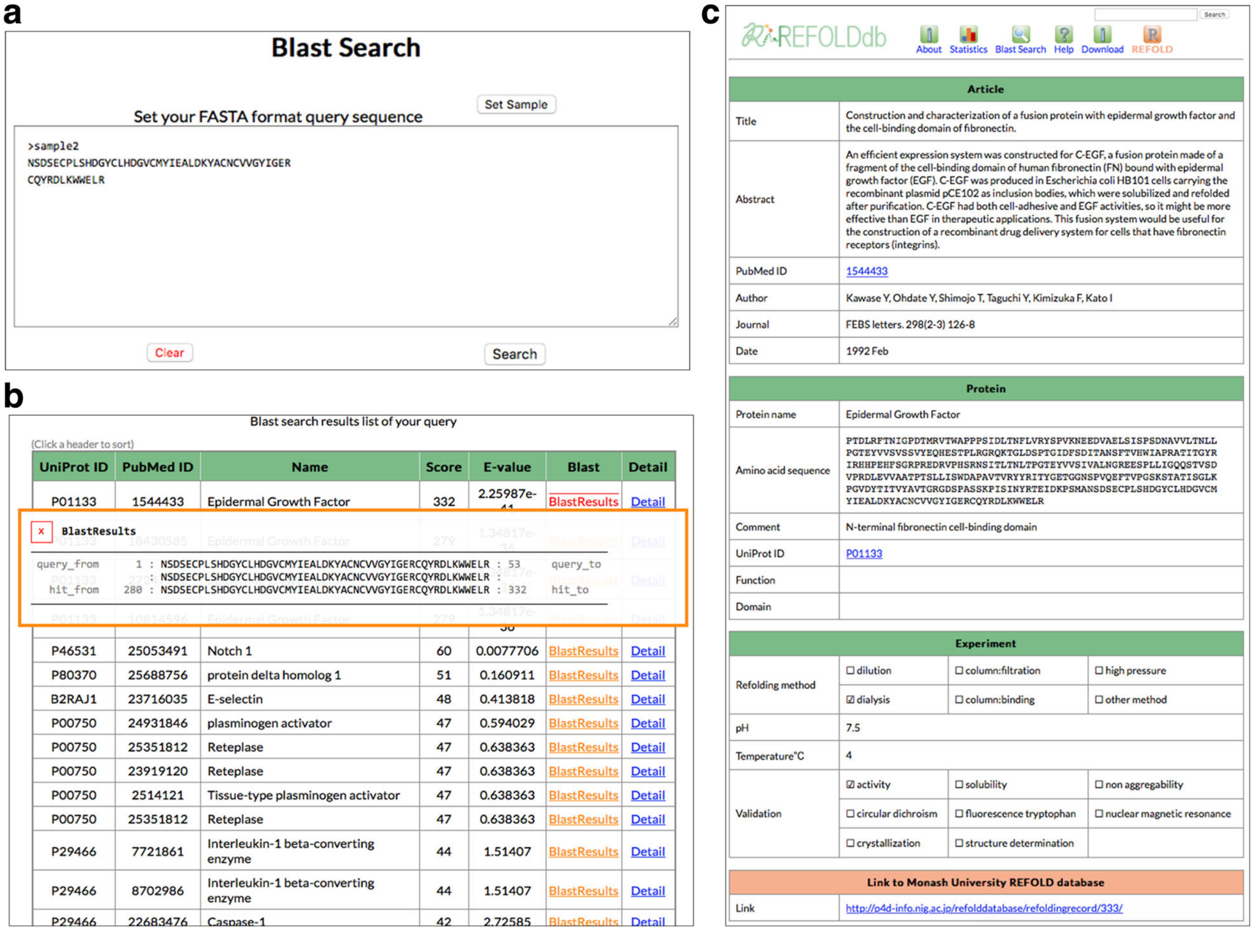
Fig. 1 The horizontal menu bar of REFOLDdb

Utility and discussion

The top page of REFOLDdb is composed of (a) a horizontal bar menu and (b) a large main search window.

(a) The horizontal bar menu
The menu includes 7 icons, "REFOLDdb", "About", "Statistics", "Blast Search", "Help", "Download" and "REFOLD", corresponding to options of database operations and 2 icons for language selection as shown in Fig. 1.

- "REFOLDdb" at the left end of the menu is a back button for the REFOLDdb user to return to the top page after several operations. "About" refers to a brief introduction of REFOLDdb and the REFOLD database. "Statistics" introduces the anatomy of REFOLDdb by graphically displaying the numbers of experiments by journals, refolding methods developed/used, pH, temperature, protein size, methods for validation, and also the number of refolding experiments by year. In addition, a "Statistics" page provides a search function, which is explained below.
- "Blast Search" page accepts a an amino acid sequence (AAseq) in FASTA format to search for "similar" proteins that were successfully refolded. A sub-menu "set sample" placed just above the blast search window toggles short, medium and long AAseqs for a quick trial (Fig. 2a). Figure 2b introduces the blast search result in a table format by choosing the medium length AAseq in the



Blast Search

Set your FASTA format query sequence

```
>sample2
NSDSECPLSHDGYCLHDGVCMYIEALDKYACNCVVGYIGER
CQYRDLKWWELR
```

Blast search results list of your query

(Click a header to sort)

UniProt ID	PubMed ID	Name	Score	E-value	Blast	Detail
P01133	1544433	Epidermal Growth Factor	332	2.25987e-41	BlastResults	Detail
P01133	18430585	Epidermal Growth Factor	279	1.34817e-36	BlastResults	Detail
P01133	2724[illegible]	[illegible]	[illegible]	1.34817e-3[illegible]	BlastResults	Detail
P01133	10814596	Epidermal Growth Factor	279	1.34817e-36	BlastResults	Detail
P46531	25053491	Notch 1	60	0.0077706	BlastResults	Detail
P80370	25688756	protein delta homolog 1	51	0.160911	BlastResults	Detail
B2RAJ1	23716035	E-selectin	48	0.413818	BlastResults	Detail
P00750	24931846	plasminogen activator	47	0.594029	BlastResults	Detail
P00750	25351812	Reteplase	47	0.638363	BlastResults	Detail
P00750	23919120	Reteplase	47	0.638363	BlastResults	Detail
P00750	2514121	Tissue-type plasminogen activator	47	0.638363	BlastResults	Detail
P00750	25351812	Reteplase	47	0.638363	BlastResults	Detail
P29466	7721861	Interleukin-1 beta-converting enzyme	44	1.51407	BlastResults	Detail
P29466	8702986	Interleukin-1 beta-converting enzyme	44	1.51407	BlastResults	Detail
P29466	22683476	Caspase-1	42	2.72585	BlastResults	Detail

```
× BlastResults

query_from     1 : NSDSECPLSHDGYCLHDGVCMYIEALDKYACNCVVGYIGERCQYRDLKWWELR : 53     query_to
                 : NSDSECPLSHDGYCLHDGVCMYIEALDKYACNCVVGYIGERCQYRDLKWWELR :
  hit_from   280 : NSDSECPLSHDGYCLHDGVCMYIEALDKYACNCVVGYIGERCQYRDLKWWELR : 332    hit_to
```

Article	
Title	Construction and characterization of a fusion protein with epidermal growth factor and the cell-binding domain of fibronectin.
Abstract	An efficient expression system was constructed for C-EGF, a fusion protein made of a fragment of the cell-binding domain of human fibronectin (FN) bound with epidermal growth factor (EGF). C-EGF was produced in Escherichia coli HB101 cells carrying the recombinant plasmid pCE102 as inclusion bodies, which were solubilized and refolded after purification. C-EGF had both cell-adhesive and EGF activities, so it might be more effective than EGF in therapeutic applications. This fusion system would be useful for the construction of a recombinant drug delivery system for cells that have fibronectin receptors (integrins).
PubMed ID	1544433
Author	Kawase Y, Ohdate Y, Shimojo T, Taguchi Y, Kimizuka F, Kato I
Journal	FEBS letters. 298(2-3) 126-8
Date	1992 Feb

Protein	
Protein name	Epidermal Growth Factor
Amino acid sequence	PTDLRFTNIGPDTMRVTWAPPPSIDLTNFLVRYSPVKNEEDVAELSISPSDNAVVLTNLL PGTEYVVSVSSVYEQHESTPLRGRQKTGLDSPTGIDFSDITANSFTVHWIAPRATITGYR IRHHPEHFSGRPREDRVPHSRNSITLTNLTPGTEYVVSIVALNGREESPLLIGQQSTVSD VPRDLEVVAATPTSLLISWDAPAVTVRYYRITYGETGGNSPVQEFTVPGSKSTATISGLK PGVDYTITVYAVTGRGDSPASSKPISINYRTEIDKPSMANSDSECPLSHDGYCLHDGVCM YIEALDKYACNCVVGYIGERCQYRDLKWWELR
Comment	N-terminal fibronectin cell-binding domain
UniProt ID	P01133
Function	
Domain	

Experiment			
Refolding method	☐ dilution	☐ column:filtration	☐ high pressure
	☑ dialysis	☐ column:binding	☐ other method
pH	7.5		
Temperature°C	4		
Validation	☑ activity	☐ solubility	☐ non aggregability
	☐ circular dichroism	☐ fluorescence tryptophan	☐ nuclear magnetic resonance
	☐ crystallization	☐ structure determination	

Link to Monash University REFOLD database	
Link	http://p4d-info.nig.ac.jp/refolddatabase/refoldingrecord/333/

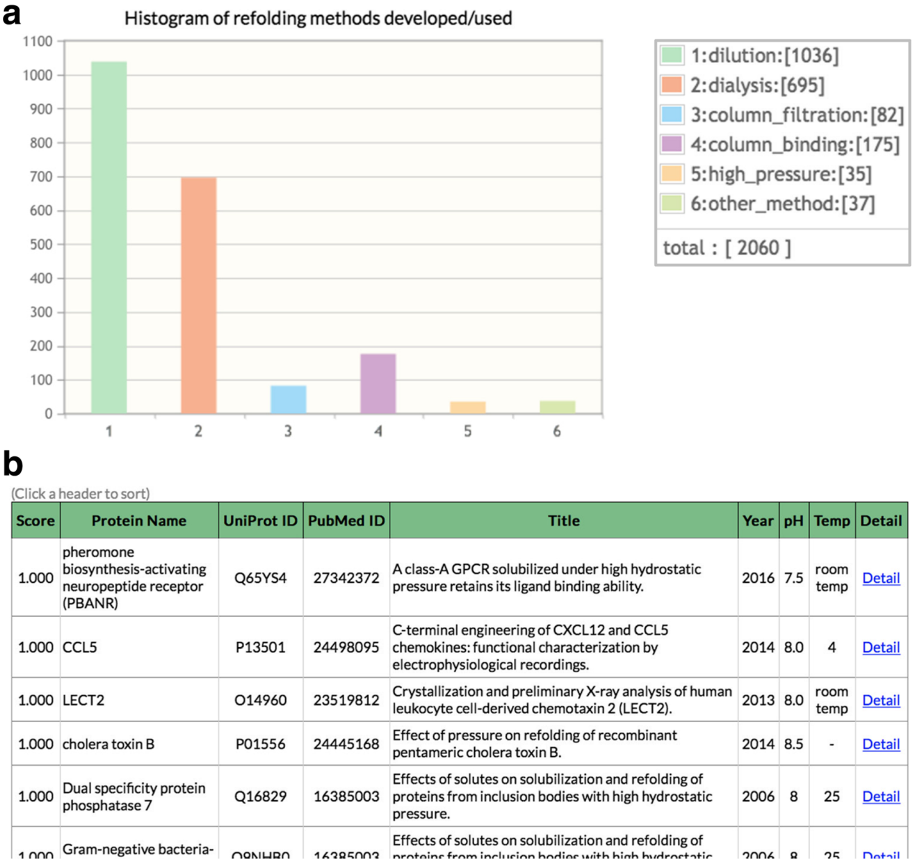
Fig. 2 Results of "Blast Search" by one of the three sample amino acids sequences (the middle length). **a**) The blast search window with a sample AAseq. **b**) The result in a table format and the alignment of AAseqs of the query and the top hit in the table. **c**) Details of the top hit

sample set. By clicking “Blast Results” in the table, the alignment between the query sequence and AAseqs in the database is displayed as shown by the box that overlaps the table. “Detail” button navigates the user to the full record as shown in Fig. 2c. The full record includes a link to the corresponding record in the REFOLD database, if available.

- “HELP”, “DOWNLOAD”, and “REFOLD” allow: browsing a compact manual for the utilization of REFOLDdb featuring screen captures, downloading the data contents of REFOLDdb in a tab-separated values (TSV) file, and accessing the previous REFOLD database respectively.

A “Statistics” page provides the user with an overview on REFOLDdb records and also a search interface. The pie and bar charts are clickable to retrieve the relevant data entries from REFOLDdb. The histogram of refolding methods developed/used is exemplified in Fig. 3a. It is obvious in the bar chart that “dilution” is the most popular methods and “high_pressure” is rarely used. It is straightforward to become familiar with “high_pressure” method by clicking the bar in Fig. 3a. The data entries that contribute to the bar are displayed in a sortable table. The top 5 ~ 65 records in the table are shown in Fig. 3b. The user is able to directly reach the full description of the method in the database using the “Detail” button in the table and then the original articles, e.g. “A class-A GPCR solubilized under high hydrostatic pressure retains its ligand binding ability” [20].

“Statistics” on experimental conditions such as pH (Fig. 4) and temperature (Fig. 5) might be useful for protein crystallographers: the histogram in Fig. 4 suggests that protein refolding experiments are most successful in a pH range of 7 to 10 regardless of other factors; the histogram in Fig. 5 shows that protein refolding experiments have been mainly performed at two temperature ranges of 0.0–4.9° Celsius (~55%) and 25.0–29.9° Celsius (~23%). We envisage that the database may allow the identification of certain refolding conditions, such as low or high temperature, which may aid downstream crystallization attempts.



(Click a header to sort)

Score	Protein Name	UniProt ID	PubMed ID	Title	Year	pH	Temp	Detail
1.000	pheromone biosynthesis-activating neuropeptide receptor (PBANR)	Q65YS4	27342372	A class-A GPCR solubilized under high hydrostatic pressure retains its ligand binding ability.	2016	7.5	room temp	Detail
1.000	CCL5	P13501	24498095	C-terminal engineering of CXCL12 and CCL5 chemokines: functional characterization by electrophysiological recordings.	2014	8.0	4	Detail
1.000	LECT2	O14960	23519812	Crystallization and preliminary X-ray analysis of human leukocyte cell-derived chemotaxin 2 (LECT2).	2013	8.0	room temp	Detail
1.000	cholera toxin B	P01556	24445168	Effect of pressure on refolding of recombinant pentameric cholera toxin B.	2014	8.5	-	Detail
1.000	Dual specificity protein phosphatase 7	Q16829	16385003	Effects of solutes on solubilization and refolding of proteins from inclusion bodies with high hydrostatic pressure.	2006	8	25	Detail
1.000	Gram-negative bacteria-	Q9NHB0	16385003	Effects of solutes on solubilization and refolding of proteins from inclusion bodies with high hydrostatic	2006	8	25	Detail

Fig. 3 Statistics for data retrieval. **a**) Histogram of refolding methods developed/used. **b**) A part of 35 REFOLDdb entries that compose the “high-pressure” bar in the Fig. 3a

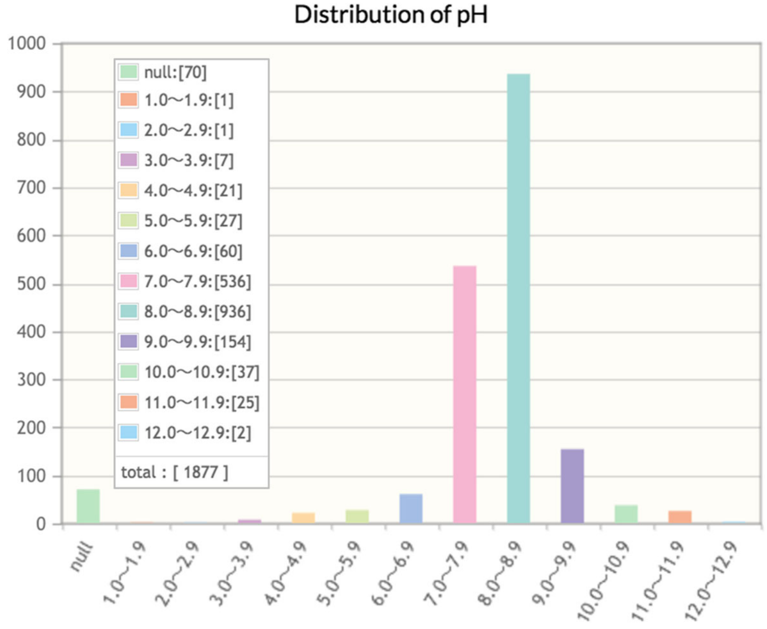


Fig. 4 Statistics on an experimental condition, pH

"Statistics" on properties of proteins could also be informative for optimal experimental design, e.g. disulfide bonds, domains, isoelectric point, metal ions, size and species. Performing this analysis shows that refolding techniques are available for a protein size range comparable to that in the PDB (Fig. 6) [21]. Both graphs imply that proteins of 100–400 amino acids are frequently analyzed. It is to be noted that REFOLDdb contains a diverse cross-section of protein architectures, including extracellular domains, subunits, whole proteins and multiple proteins. It is possible in theory to transform statistics to an inference engine based on AAseqs. However, it requires stringent cleansing of AAseqs in research articles and databases

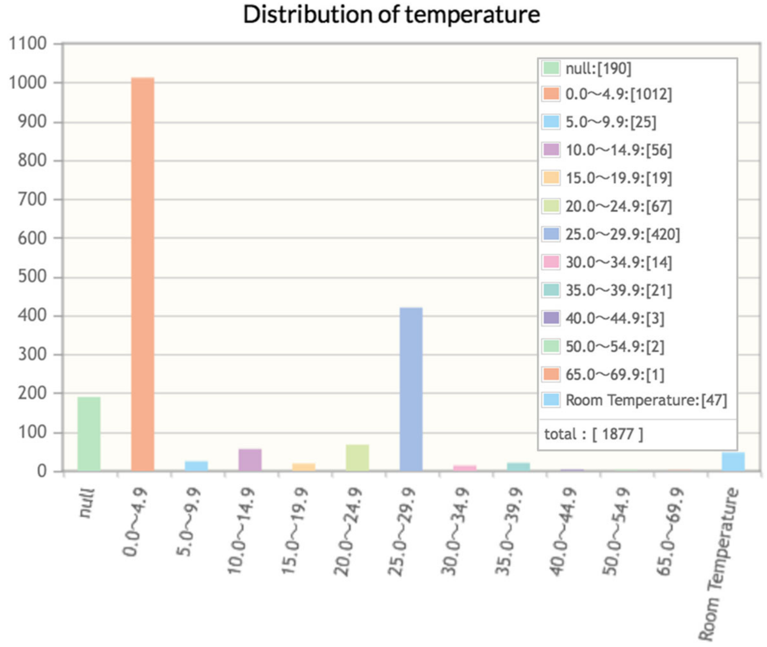


Fig. 5 Statistics on an experimental condition, temperature

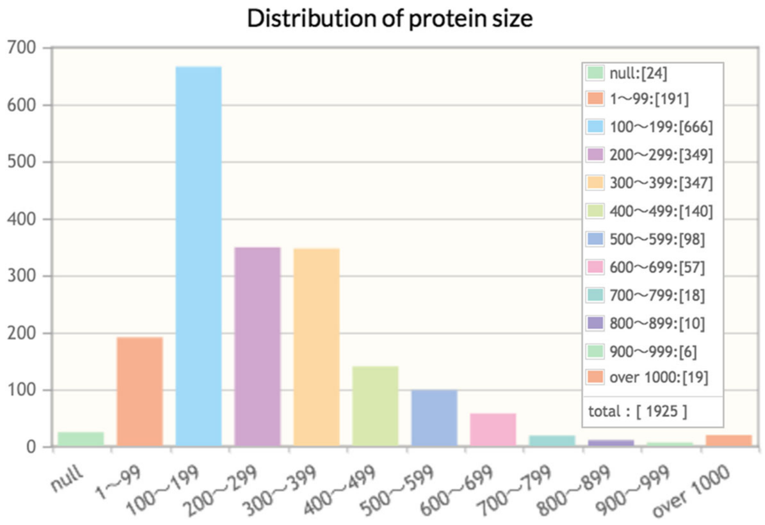
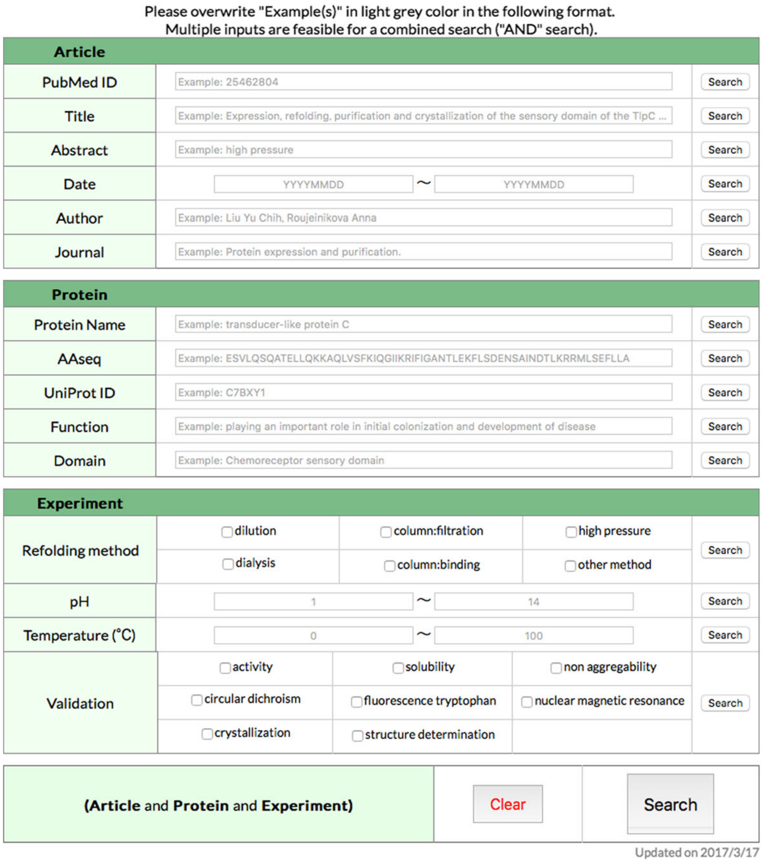


Fig. 6 Statistics on size of proteins refolded



Please overwrite "Example(s)" in light grey color in the following format.
Multiple inputs are feasible for a combined search ("AND" search).

Article		
PubMed ID	Example: 25462804	Search
Title	Example: Expression, refolding, purification and crystallization of the sensory domain of the TlpC ...	Search
Abstract	Example: high pressure	Search
Date	YYYYMMDD ～ YYYYMMDD	Search
Author	Example: Liu Yu Chih, Roujeinikova Anna	Search
Journal	Example: Protein expression and purification.	Search

Protein		
Protein Name	Example: transducer-like protein C	Search
AAseq	Example: ESVLQSQATELLQKKAQLVSFKIQGIIKRIFIGANTLEKFLSDENSAINDTLKRRMLSEFLLA	Search
UniProt ID	Example: C7BXY1	Search
Function	Example: playing an important role in initial colonization and development of disease	Search
Domain	Example: Chemoreceptor sensory domain	Search

Experiment				
Refolding method	□dilution □dialysis	□column:filtration □column:binding	□high pressure □other method	Search
pH	1 ～ 14			Search
Temperature (°C)	0 ～ 100			Search
Validation	□activity □circular dichroism □crystallization	□solubility □fluorescence tryptophan □structure determination	□non aggregability □nuclear magnetic resonance	Search

(Article and **Protein** and **Experiment)**	Clear	Search

Updated on 2017/3/17

Fig. 7 The main window for searching REFOLDdb

that often implicitly include tags and linkers. It is also a difficult task to collect negative data that are prerequisite for the development of a reliable inference engine. Nevertheless, REFOLDdb is a good starting point for data mining in order to customize experimental conditions for a given protein in the future.

(b) The main search window (Fig. 7).
The window is located just under the horizontal bar menu. A combined search of REFOLDdb can be carried out in the following two steps:

1) Overwrite “Example(s)” in light grey color and/or check boxes as many as needed. In the case of the data items of “pH” and “Temperature”, lower limit and/or upper limit can be specified.
2) Click a “Search” button in the line of a data item to go through the specified data item, or click the “Search” button at the bottom-right corner of the search window to perform a combined search, namely, “AND” search of multiple data items.

Multiple hits to a query will be displayed in a table format that is composed of sortable columns of “Protein Name”, “UniProt ID”, “PubMed ID”/”Title”/”Year” of the publication, “pH”, “Temp(erature)”. “Detail” buttons in the table navigates the user to detailed information on proteins and experimental conditions.

Conclusions

The resources, including human resources, required for running and updating REFOLDdb is kept to a minimum. A team of one annotator who is knowledgeable about structural biology and a part time system engineer will be able to keep REFOLDdb up-to-date as far as collecting data from research papers on a monthly basis. The database system based on the virtual machine is almost autonomous and also flexible enough to allow future expansion.

In the future, we will evaluate new data sources other than research articles, such as patents, that might make the database more comprehensive. In addition, we will investigate the implementation of data-mining functionality to allow the prediction of suitable refolding methods based on chemical, physical and/or genetic features of proteins that have been successfully refolded.

Abbreviations

REFOLD database: The database on refolding technologies developed by Monash University; REFOLDdb: The database on refolding technologies developed by the authors; Tech_db: Databases on technologies

Acknowledgements

The authors are grateful to Professor Junichi Takagi (Institute for Protein Research, Osaka University) for his suggestions on the needs of the research communities to databases in structural life sciences.

Funding

This work was supported by the ‘Platform for Drug Discovery, Informatics, and Structural Life Science’ grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) and the Japan Agency for Medical Research and Development (AMED). The funding body did not play any role in the design or conclusion of the study. Funding for open access charge: Waseda University, Japan.

Availability of data and materials

REOLDdb is searchable and downloadable at http://p4d-info.nig.ac.jp/refolddb/. A list of all papers which REFOLDdb referred to is accessible from the menu “List of publications referred by REFOLDdb” in “About” page at http://p4d-info.nig.ac.jp/refolddb/about.cgi?lang=EN.

Authors’ contributions

HM, HS, TS, JO, KN and MT designed the REFOLD db. HM, KM, JO, TS, SN, YX, DW and HU contributed to the production of the data set based on research papers on refolding technologies. AB contributed to the conversion of the REFOLD database to REFOLD db. HS, KN, AB and KY wrote the manuscript. All authors read and approved the final manuscript.

Authors’ information

Not applicable.

Competing interests

Not applicable.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Center for Information Biology, National Institute of Genetics, 1111 Yata Mishima, Shizuoka 411-8540, Japan. [2]The Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia. [3]Laboratory of Protein Synthesis and Expression, Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan. [4]Laboratory of Structural Biology and Food Biotechnology, Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan. [5]Graduate School of Humanities and Sciences, Ochanomizu University, 2-1-1 Otsuka, Bunkyo, Tokyo 112-8610, Japan. [6]Center for Simulation Science and Informational Biology, Ochanomizu University, 2-1-1 Otsuka, Bunkyo, Tokyo 112-8610, Japan. [7]School of Advanced Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjyuku, Tokyo 169-8555, Japan.

Received: 8 December 2016 Accepted: 11 April 2017
Published online: 24 April 2017

References

1. Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microb Cell Fact. 2015;14:41.
2. Eiberle MK, Jungbauer A. Technical refolding of proteins: Do we have freedom to operate? Biotechnol J. 2010;5(6):547–59.
3. Yamaguchi H, Miyazaki M. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. Biomolecules. 2014;4(1):235–51.
4. Li M, Su ZG, Janson JC. In vitro protein refolding by chromatographic procedures. Protein Expr Purif. 2004;33(1):1–10.
5. Okai M, Ohtsuka J, Asano A, Guo L, Miyakawa T, Miyazono K, Nakamura A, Okada A, Zheng H, Kimura K, Nagata K, Tanokura M. High pressure refolding, purification, and crystallization of flavin reductase from Sulfolobus tokodaii strain 7. Protein Expr Purif. 2012;84(2):214–8.

6. Coutard B, Danchin EG, Oubelaid R, Canard B, Bignon C. Single pH buffer refolding screen for protein from inclusion bodies. Protein Expr Purif. 2012;82(2):352–9.
7. Xie Y, Wetlaufer DB. Control of aggregation in protein refolding: the temperature-leap tactic. Protein Sci. 1996;5(3):517–23.
8. Gupta P, Hall CK, Voegler AC. Effect of denaturant and protein concentrations upon protein refolding and aggregation: a simple lattice model. Protein Sci. 1998;7(12):2642–52.
9. Yamaguchi S, Yamamoto E, Mannen T, Nagamune T, Nagamune T. Protein refolding using chemical refolding additives. Biotechnol J. 2013;8(1):17–31.
10. Phan J, Yamout N, Schmidberger J, Bottomley SP, Buckle AM. Refolding your protein with a little help from REFOLD. Methods Mol Biol. 2011;752:45–57.
11. Buckle AM, Devlin GL, Jodun RA, Fulton KF, Faux N, Whisstock JC, Bottomley SP. The matrix refolded. Nat Methods. 2005;2(1):3.
12. Chow MK, Amin AA, Fulton KF, Fernando T, Kamau L, Batty C, Louca L, Ho S, Whisstock JC, Bottomley SP, Buckle AM. The REFOLD database: a tool for the optimization of protein expression and refolding. Nucleic Acids Res. 2006;34(Database issue):D207–12.
13. Chow MK, Amin AA, Fulton KF, Whisstock JC, Buckle AM, Bottomley SP. REFOLD: an analytical database of protein refolding methods. Protein Expr Purif. 2006;46(1):166–71. Epub 2005 Aug 15.
14. Editorial. Database under maintenance. Nature Methods. 2016;13(9):699
15. Oliver SG, Lock A, Harris MA, Nurse P, Wood V. Model organism databases: essential resources that need the support of both funders and users. BMC Biology. 2016;14:49.
16. International Nucleotide Sequence Database. http://www.insdc.org/ Accessed 2 Dec 2016.
17. Worldwide Protein Data Bank (wwPDB). http://www.wwpdb.org/ Accessed 2 Dec 2016.
18. UniProt. http://www.uniprot.org/ Accessed 2 Dec 2016.
19. Oates ME, Stahlhacke J, Vavoulis DV, Smithers B, Rackham OJ, Sardar AJ, Zaucha J, Thurlby N, Fang H, Gough J. The SUPERFAMILY 1.75 database in 2014: a doubling of data. Nucleic Acids Res. 2015;43(Database issue):D227–33.
20. Katayama Y, Suzuki T, Ebisawa T, Ohtsuka J, Wang S, Natsume R, Lo YH, Senda T, Nagamine T, Hull JJ, Matsumoto S, Nagasawa H, Nagata K, Tanokura T. A class-A GPCR solubilized under high hydrostatic pressure retains its ligand binding ability. Biochim Biophys Acta. 2016;1858(9):2145–51.
21. RCSB Protein Data Bank. Residue Count Histogram http://www.rcsb.org/pdb/static.do?p=general_information/pdb_statistics/index.html Accessed 28 Feb 2017.