Data in Brief 8 (2016) 948–957

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data on publications, structural analyses, and queries used to build and utilize the AlloRep database

Filipa L. Sousa [a,1], Daniel J. Parente [b,2], Jacob A. Hessman [b,3], Allen Chazelle [b,4], Sarah A. Teichmann [c,d], Liskin Swint-Kruse [b,*]

[a] *Institute of Molecular Evolution, Heinrich-Heine Universität Düsseldorf, Universitätstrasse 1, 40225 Düsseldorf, Germany*
[b] *The Department of Biochemistry and Molecular Biology, The University of Kansas Medical Center, Kansas City, KS 66160, USA*
[c] *European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK*
[d] *Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*

ARTICLE INFO

Article history:
Received 3 May 2016
Received in revised form
22 June 2016
Accepted 4 July 2016
Available online 9 July 2016

ABSTRACT

The AlloRep database (www.AlloRep.org) (Sousa et al., 2016) [1] compiles extensive sequence, mutagenesis, and structural information for the LacI/GalR family of transcription regulators. Sequence alignments are presented for >3000 proteins in 45 paralog subfamilies and as a subsampled alignment of the whole family. Phenotypic and biochemical data on almost 6000 mutants have been compiled from an exhaustive search of the literature; citations for these data are included herein. These data include information about oligomerization state, stability, DNA binding and allosteric regulation. Protein structural data for 65 proteins are presented as easily-accessible, residue-contact networks. Finally, this article includes example queries to enable the use of the AlloRep database. See the related article, “AlloRep: a repository of

---

DOI of original article: http://dx.doi.org/10.1016/j.jmb.2015.09.015

[*] Corresponding author.
E-mail address: lswint-kruse@kumc.edu (L. Swint-Kruse).

[1] Current address: Department of Ecogenomics and Systems Biology, University of Vienna, Vienna, Austria.

[2] Current address: Department of Family Medicine, The University of Kansas School of Medicine.

[3] Current address: Department of Pediatrics, The University of Kansas School of Medicine.

[4] Current address: Department of Clinical Laboratory Sciences, The University of Kansas School of Medicine.

http://dx.doi.org/10.1016/j.dib.2016.07.006

sequence, structural and mutagenesis data for the LacI/GalR transcription regulators" (Sousa et al., 2016) [1].

Specifications Table

Subject area	*Biology*
More specific subject area	*Protein biochemistry*
Type of data	*Text, figure*
How data was acquired	*Literature survey and computational calculations for LacI/GalR protein variants*
Data format	*Normalized; analyzed*
Experimental factors	*Mutational data were normalized to wild-type protein activity*
Experimental features	*For structural data, intra- and intermolecular non-covalent contacts were calculated at a 5A threshold.*
Data source location	*The University of Kansas Medical Center, Kansas City, KS*
Data accessibility	*Data is within this article and available at* www.AlloRep.org

Value of the data

- The AlloRep database (www.AlloRep.org) compiles extensive sequence, mutagenesis, and structural information for the LacI/GalR family of transcription regulators.
- The AlloRep database simplifies the consolidation of non-covalent structural information with mutagenesis and sequence conservation data.
- The AlloRep database can be used to benchmark computational predictions and to design synthetic transcription repressors for biotechnology.
- The example queries contained in this article can be used to improve searches of the AlloRep database.

1. Data

The AlloRep database (www.AlloRep.org) [1] compiles extensive sequence, mutagenesis, and structural information for the LacI/GalR family of transcription regulators. Phenotypic and biochemical data on almost 6000 mutants have been compiled from an exhaustive search of the literature; citations for these data are listed in this publication [2–82]. The data can be exported to build a local copy on the user's computer, but the insert and import features are disabled. New data are welcome and can be submitted to the corresponding author at lswint-kruse@kumc.edu. Here, we detail the organization of the 5 database modules and their components tables, and provide full descriptions for the contents of table columns. Fig. 1 overviews the structure of the database.

We also present a protein structural comparison that was facilitated by compiling the information in the structural module. Fig. 2 shows a comparison of intra- and inter-molecular contacts from a comparative study of 65 structures available for the LacI/GalR homologs.