Data Article

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

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

a Institute of Molecular Evolution, Heinrich-Heine Universität Düsseldorf, Universitätsstrasse 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.
sequence, structural and mutagenesis data for the LacI/GalR transcription regulators” (Sousa et al., 2016) [1].
© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

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 5Å 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.
Finally, the database can be searched by selecting a table from one of the modules and using the built-in search fields (search tab; Fig. 3). In addition, command line queries can be executed using the SQL tab. Example command line queries are listed in supplement to this manuscript.

2. Experimental design, materials and methods

2.1. AlloRep database overview and description of modules

The AlloRep database, freely available at www.AlloRep.org, [1] is divided into five modules (Fig. 1). Below are explanations of relevant tables and abbreviation used in each section. The tables can be browsed within the website and sorted by clicking on various column headings. In addition, example command line SQL queries are given in the supplement that can be used to link the information between the various modules.

2.2. Module 1: mutagenesis data

This module contains information collected from an exhaustive literature search [2–82]. The module entails two tables: “mut1_single” and “mut2_combinatorial”. For variants in “mut1_single”, all outcomes can be attributed to a single mutation, either by comparing the properties of a single
Fig. 2. Comparisons of inter- and intra-molecular contacts among 65 structures of LacI/GalR homologs. All available structures were collected for each equivalent state of a given protein (from the same species and bound the same ligands), including the occurrences of multiple structures present in a unit cell. Inter- and intra-monomeric contacts were determined as defined in the text, and the frequency of each contact was calculated for the set of structures. If only one structure was available, the frequency was set to 100% by default. As an example, panel (a) shows an excerpt from a matrix containing information about the frequency of various contacts for all structures of *E.coli* apo-PurR. Each contact matrix was then linearized in numerical order (b) to make one column of panel (c). As a second example, the dashed box contains the composite information for all structures of LacI bound to DNA and the small molecule NPF. In panel (c), the contacts were ordered on the Y axis so that those involving the N-terminal DNA binding domain are at the top, those of the linker come next (positions 45–62 in *E. coli* LacI), followed by contacts in the regulatory domain. Each column along the X axis corresponds to the named group of equivalent structures. Bound ligands are in parentheses and ligand abbreviations can be found in the table "struct2_ligand_description." Different colors indicate the frequency that a particular contact occurs. Inter-monomeric contacts are collected on the left of panel (c). Some structures contained monomers that could not be dimerized by symmetry operations; thus their inter-monomer contacts could not be determined. Intra-monomeric contacts are shown on the right. Once contact frequency was calculated, agnostic, hierarchical clustering was used to order the inter- and intra-monomeric contacts in panel (c). These plots show that the inter-monomer contacts (left panel) cluster according to their ligand binding state. For example, the DNA bound structures for different homologs are more similar to each other than to their respective inducer bound structures. In contrast, the intra-monomeric contacts (right panel) cluster so that the structures for each LacI/GalR subfamily are most closely related, regardless of their binding state.
mutation to those of the wild-type protein, or, for example, by comparing a double mutant to a variant that contains the relevant single mutation. Variants in the “mut2_combinatorial" table contain multiple mutations that have not yet been parsed into their component contributions. Both tables contain fields for: a unique internal_id for each variant, the relevant LacI/GalR subfamily, species of origin, position number in the parent protein, position number translated to the LacI reference numbering system, one-letter codes for the original amino acid and the mutational variant, and PMIDs of the original publications. The mut1_single also contains the parent protein that provides the basis for comparison of experimental results.

In both tables, additional columns contain all available experimental information for the variant. Since experiments were carried out over several decades, in different laboratories, and sometimes under different experimental conditions (such as different buffers), the functional effects of each mutation are reported relative to the appropriate parent protein. Information regarding the effect on protein secondary structure and/or oligomerization state (where "D" stands for dimer, "T" for tetramer and "M" for monomer) are stored in columns with those names. Effects on urea stability, thermal denaturation, trypsin digestion assays, and temperature sensitivity are stored in other columns. The phenotypic and biochemical characterizations are provided in the “phenotype", “allostery" and “reverse phenotype" columns. When possible, the relative differences are indicated with the symbols: [0] or [−−] for total loss, [−−] for a significant decrease, [−] small decrease, [=] or ~ if comparable with wild type, [+] for small increase and [+++] for a significant increase. Any additional information is provided in the “observation" column.

2.3. Module 2: sequence data

This module contains three tables with: (i) the manually-curated alignment of representative sequences for the entire LacI/GalR family (each homolog is contained in a separate row) [83]; (ii) the separate alignments for all subfamilies (each subfamily alignment is contained in one row); and (iii) a
table containing unaligned “orphan” sequences (one per row) that do not match any of the current subfamilies. All data are stored in fasta format. After selecting a table of interest, it can be downloaded using the export button at the bottom of the page and selecting the desired format. Note that the output options can be customized for a better compatibility with the user’s operating system. The subfamily alignments can be matched to the spacing of the manually-curated, whole-family alignment using the program MARS-Prot (https://github.com/djparente/MARS) [84].

2.4. Module 3: structural data

All available structures for LacI/GalR homologs [18,21,55,57,66,85–102] were retrieved from the Protein Data Bank database [103]. This module contains all the information regarding the PDB description (struct1_pdb_overview table), available ligand information (struct2_ligand_description table), and four tables with different types of contacts.

For each LacI/GalR structure, non-covalent contacts were defined when any two residues had at least one non-hydrogen atom within 5 Å of the other. Angles and other geometries were not considered. For all structures, the full set of contacts is stored in the table “struct3_contacts_monomers” where contacts were grouped according to their protein subfamily, inter- or intra-monomeric nature, and ligand. Next, for the table “struct4_contacts_heatmap”, equivalent structures (those for the same protein and liganded state) were combined to calculate the frequency of each contact pair; these values are presented in a single column for each group of equivalent structures (Fig. 2). For example, apo LacI has two structures (1lbi and 3edc) each of which contains four monomers. In two of the 8 chains (25%), LacI residues E100 and C107 are within 5 Å of each other; thus the occupancy score for this contact is 25%. For states that have only one available structure, the default value is 100%.

The table “struct5_contacts_macromol” contains information regarding the contacts between the LacI/GalR proteins and macromolecular ligands such as DNA or heteroproteins. Contacts between LacI/GalR proteins and small-molecule ligands are stored in the table “struct6_contacts_ligand table”, which also includes information on the total contact surface area and the number of contacts.

2.5. Module 4: translation tables

This section contains two tables – “translate_numbering_table” and “translate_numbers_to_laci” – that allow the conversion between the numbering system of Escherichia coli LacI and those of other LacI/GalR homologs. “Translate_numbers_to_laci” contains the necessary information for connecting both structural or mutagenesis data to the “translate_numbering_table”. The “translate_numbering_table” contains the structural alignment of all crystallographic structures as well as representative sequences for each protein subfamily that has available mutagenesis data.

Using either the PDB identifier and residue numbering as input (from tables in the structural module) or information regarding the LacI/GalR subfamily and residue numbering as input (from tables in the mutation module), the user can obtain the code to be used in the translation_numbers_to_laci and retrieve the original sequence numbering.

2.6. Module 5: citations

A final module comprises one table (“x_data_sources_cited”) that contains all bibliographic information and can be queried using the PMID or the citation code provided in the structural and mutagenesis tables.

Acknowledgments

This work was supported by Fundação para a Ciência e Tecnologia, SFRH/BPD/73058/2010 (FLS), NIH GM 079423 (LSK), and the University of Kansas Medical Center Biomedical Research Training Program (DJP). We thank Tina Perica for many stimulating discussions about this project.
Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.07.006.

References

[1] F.L. Sousa, D.J. Parente, D.L. Shis, J.A. Hessman, A. Chazelle, M.R. Bennett, et al., AlloRep: a repository of sequence, structural and mutagenesis data for the LacI/GalR transcription regulators, J. Mol. Biol. 428 (2016) 671–678.

[2] P. Markiewicz, L.G. Kleina, C. Cruz, S. Ehet, J.H. Miller, Genetic studies of the lac repressor. XIV. Analysis of 4000 altered E. coli lac repressors reveals essential and non-essential residues, as well as “spacers” which do not require a specific sequence, J. Mol. Biol. 240 (1994) 421–433.

[3] J.H. Miller, Genetic studies of the lac repressor. XII. Amino acid replacements in the DNA binding domain of the E. coli lac repressor, J. Mol. Biol. 180 (1984) 205–212.

[4] L.G. Kleina, J.H. Miller, Genetic studies of the lac repressor. XIII. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors, J. Mol. Biol. 212 (1990) 295–318.

[5] J. Suckow, P. Markiewicz, L.G. Kleina, J. Miller, B. Kisters-Woike, B. Müller-Hill, Genetic studies of the lac repressor. XV. 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure, J. Mol. Biol. 261 (1996) 509–523.

[6] A.E. Chakerian, V.M. Tesmer, S.P. Manly, J.K.Brackett, M.J. Lynch, J.T. Hoh, et al., Evidence for leucine zipper motif in lactose repressor protein, J. Biol. Chem. 266 (1991) 1371–1374.

[7] A. Kraus, E. Kuster, A. Wagner, K. Hoffmann, W. Hillen, Identification of a co-repressor binding site in catabolite control protein CcpA, Mol. Microbiol. 30 (1998) 955–963.

[8] B.H. Kallipolitis, P. Valentin-Hansen, A role for the interdomain linker region of the E. coli CytR regulator in repression complex formation, J. Mol. Biol. 342 (2004) 1–7.

[9] J.H. Miller, C. Coulondre, M. Hofer, U. Schmeissner, H. Sommer, A. Schmitz, et al., Genetic studies of the lac repressor. IX. Generation of altered proteins by the suppression of nonsense mutations, J. Mol. Biol. 131 (1979) 191–222.

[10] M. Geanacopoulos, S. Adhya, Genetic analysis of GalR tetramerization in DNA looping during repressosome assembly, J. Biol. Chem. 277 (2002) 33148–33152.

[11] J.H. Miller, C. Coulondre, M. Hofer, U. Schmeissner, H. Sommer, A. Schmitz, et al., Genetic studies of the lac repressor. X. Generation of altered proteins by the suppression of nonsense mutations, J. Mol. Biol. 131 (1979) 261–266.

[12] M. Geanacopoulos, G. Vasmatzis, V.B. Zhurkin, S. Adhya, Gal repressosome contains an antiparallel DNA loop, Nat. Struct. Biol. 8 (2001) 432–436.

[13] A.M. Khoury, H.S. Nick, P. Lu, in vivo interaction of Bacillus megaterium catabolite control protein CcpA mutants affected in the DNA- and cofactor-binding domains, Microbiology 150 (2004) 613–620.

[14] S. Meyer, R. Ramot, K. Kishore Inampudi, B. Luo, C. Lin, S. Amere, et al., Engineering alternate cooperative-communications in the lactose repressor protein scaffold, Protein Eng. Des. Sel. 26 (2013) 433–443.

[15] F. Lu, M.A. Schumacher, D.N. Arvidson, A. Haldimann, B.L. Wanner, H. Zalkin, et al., Structure-based redesign of corepressor specificity of the E. coli purine repressor by substitution of residue 190, Biochemistry 37 (1998) 971–982.

[16] W.Y. Chou, K.S. Matthews, Mutation in hinge region of lactose repressor protein alters physical and functional properties, J. Biol. Chem. 264 (1989) 6171–6176.

[17] M. Geanacopoulos, V. Monedero, H. Zalkin, S. Adhya, Gal repressosome contains an antiparallel DNA loop, Nat. Struct. Biol. 8 (2001) 432–436.

[18] C.D. Esteban, K. Mahr, V. Monedero, W. Hillen, G. Perez-Martinez, F. Titgemeyer, Complementation of a Delta ccpA mutant reveals essential and non-essential residues, as well as “spacers” which do not require a specific sequence, J. Mol. Biol. 240 (1994) 421–433.

[19] E. Kuster-Schock, A. Wagner, U. Volker, W. Hillen, Mutations in catabolite control protein CcpA showing glucose-independent regulation in Bacillus megaterium, J. Bacteriol. 181 (1999) 4125–4128.

[20] C.D. Esteban, K. Mahr, V. Monedero, W. Hillen, G. Perez-Martinez, F. Titgemeyer, Complementation of a Delta ccpA mutant of Lactobacillus casei with CcpA mutants affected in the DNA- and cofactor-binding domains, Microbiology 150 (2004) 513–523.

[21] C.D. Esteban, K. Mahr, V. Monedero, W. Hillen, G. Perez-Martinez, F. Titgemeyer, Complementation of a Delta ccpA mutant of Lactobacillus casei with CcpA mutants affected in the DNA- and cofactor-binding domains, Microbiology 150 (2004) 513–523.

[22] C.D. Esteban, K. Mahr, V. Monedero, W. Hillen, G. Perez-Martinez, F. Titgemeyer, Complementation of a Delta ccpA mutant of Lactobacillus casei with CcpA mutants affected in the DNA- and cofactor-binding domains, Microbiology 150 (2004) 513–523.

[23] C.D. Esteban, K. Mahr, V. Monedero, W. Hillen, G. Perez-Martinez, F. Titgemeyer, Complementation of a Delta ccpA mutant of Lactobacillus casei with CcpA mutants affected in the DNA- and cofactor-binding domains, Microbiology 150 (2004) 513–523.

[24] C.D. Esteban, K. Mahr, V. Monedero, W. Hillen, G. Perez-Martinez, F. Titgemeyer, Complementation of a Delta ccpA mutant of Lactobacillus casei with CcpA mutants affected in the DNA- and cofactor-binding domains, Microbiology 150 (2004) 513–523.
[25] L. Swint-Kruse, H. Zhan, B.M. Fairbanks, A. Maheshwari, K.S. Matthews, Perturbation from a distance: mutations that alter LacI function through long-range effects, Biochemistry 42 (2003) 14004–14016.

[26] J.K. Barry, K.S. Matthews, Substitutions at histidine 74 and aspartate 278 alter ligand binding and allosterie in lactose repressor protein, Biochemistry 38 (1999) 3579–3590.

[27] S. Semsey, M. Geanacopoulou, D.E.A. Lewis, S. Adhya, Operator-bound GalR dimers close DNA loops by direct interaction: tetramerization and inducer binding, EMBO J. 21 (2002) 4349–4356.

[28] C.S. Barbier, S.A. Short, Amino acid substitutions in the CytR repressor which alter its capacity to regulate gene expression, J. Bacteriol. 174 (1992) 2881–2880.

[29] S. Miura-Onai, M. Yabuta, K. Ohsuye, Mutational study at Ser300 position of the lac repressor protein, J. Biol. Chem. 270 (1995) 10640–10649.

[30] S. Roy, S. Semsey, M. Liu, G.N. Gussin, S. Adhya, GalR represses galP1 by inhibiting the rate-determining open complex formation through RNA polymerase contact: a GalR negative control mutant, J. Mol. Biol. 344 (2004) 609–618.

[31] H. Zhan, Z. Sun, K.S. Matthews, Functional impact of polar and acidic substitutions in the lactose repressor hydrophobic monomerinterface with a buried lysine, Biochemistry 48 (2009) 1305–1314.

[32] A.M. Mikulskis, A. Aristarkhov, E.C. Lin, Regulation of expression of the ethanol dehydrogenase gene (adhe) in E. coli by catabolite repressor activator protein Cra, J. Bacteriol. 179 (1997) 7129–7134.

[33] R.O. Spotts, A.E. Chakerian, K.S. Matthews, Arginine 197 of lac repressor contributes significantly to inducer binding, Proc. Natl. Acad. Sci. USA 109 (2012) 11336–11341.

[34] A. Glasfeld, A.N. Koehler, M.A. Schumacher, R.G. Brennan, The role of lysine 55 in determining the specificity of the purine repressor for its operators through minor groove interactions, J. Mol. Biol. 291 (1999) 347–361.

[35] J. Xu, S. Liu, M. Chen, J. Ma, K.S. Matthews, Altering residues N125 and D149 impacts sugar effector binding and allosteric parameters in E. coli lacI repressor protein, Biochemistry 50 (2011) 9002–9013.

[36] D.N. Arvidson, F. Lu, C. Faber, H. Zalkin, R.G. Brennan, The structure of PurR mutant L54M shows an alternative route to DNA kinking, Nat. Struct. Biol. 5 (1998) 436–441.
[58] J.A. Gardner, K.S. Matthews, Characterization of two mutant lactose repressor proteins containing single tryptophans, J. Biol. Chem. 265 (1990) 21061–21067.

[59] S. Meinhardt, L. Swint-Kruse, Experimental identification of specificity determinants in the domain linker of a LacI/GalR protein: bioinformatics-based predictions generate true positives and false negatives, Proteins 73 (2008) 941–957.

[60] S. Tungtir, S. Meinhardt, L. Swint-Kruse, Comparing the functional roles of nonconserved sequence positions in homologous transcription repressors: implications for sequence/function analyses, J. Mol. Biol. 395 (2010) 785–802.

[61] S. Meinhardt, M.W. Manley, D.J. Parente, L. Swint-Kruse, Rheostats and toggle switches for modulating protein function, PLoS One 8 (2013) e53502.

[62] L. Swint-Kruse, C.A. Royer, J.A. Gardner, J.M. Beechem, J.C. Brochon, K.S. Matthews, Resolution of the crystal structure of a 1.6-hexanediol bound tetrameric form of L. Swint-Kruse, H. Zhan, K.S. Matthews, Integrated insights from simulation, experiment, and mutational analysis yield new details of Lac function, Biochemistry 44 (2005) 11201–11213.

[63] J.C. Nichols, K.S. Matthews, Combinatorial mutations of lac repressor. Stability of monomer-monomer interface is increased by apolar substitution at position 94, New Biol. 3 (1991) 57–62.

[64] C.A. Royer, J.A. Gardner, J.M. Beechem, J.C. Brochon, K.S. Matthews, Resolution of the fluorescence decay of the two tryptophan residues of lac repressor using single tryptophan mutants, Biophys. J. 58 (1990) 363–378.

[65] J.K. Barry, K.S. Matthews, Ligand-induced conformational changes in lactose repressor: a fluorescence study of single tryptophan mutants, Biochemistry 36 (1997) 15632–15642.

[66] M.A. Schumacher, G.S. Allen, M. Diel, G. Seidel, W. Hillen, R.G. Brennan, Structural basis for allosteric control of the transcription regulator CcpA from the phosphoprotein HPr-Ser46-P, Cell 118 (2004) 731–741.

[67] M. Geanacopoulos, G. Vasmatzis, D.E. Lewis, S. Roy, S. Adhya, GalR mutants defective in repressosome formation, Nucleic Acids Res. 39 (2011) 2931–2942.

[68] A. Barker, R. Fickert, S. Oehler, B. Müller-Hill, Operator search by mutant lac repressor, J. Mol. Biol. 278 (1998) 549–564.

[69] S. Meinhardt, L. Swint-Kruse, Experimental identification of specificity determinants in the domain linker of a LacI/GalR protein: bioinformatics-based predictions generate true positives and false negatives, Proteins 73 (2008) 941–957.

[70] J.A. Gardner, K.S. Matthews, Characterization of two mutant lactose repressor proteins containing single tryptophans, J. Biol. Chem. 265 (1990) 21061–21067.

[71] J.C. Nichols, K.S. Matthews, Combinatorial mutations of lac repressor. Stability of monomer-monomer interface is increased by apolar substitution at position 94, New Biol. 3 (1991) 57–62.

[72] C.A. Royer, J.A. Gardner, J.M. Beechem, J.C. Brochon, K.S. Matthews, Resolution of the fluorescence decay of the two tryptophan residues of lac repressor using single tryptophan mutants, Biophys. J. 58 (1990) 363–378.

[73] J.K. Barry, K.S. Matthews, Ligand-induced conformational changes in lactose repressor: a fluorescence study of single tryptophan mutants, Biochemistry 36 (1997) 15632–15642.

[74] A. Ozarski, J.K. Barry, K.S. Matthews, A.H. Maki, Ligand-induced conformational changes in lactose repressor: a phosphorescence and ODMR study of single-tryptophan mutants, Biochemistry 38 (1999) 6715–6722.

[75] F. Dong, S. Spott, O. Zimmermann, B. Kisters-Woike, B. Müller-Hill, A. Barker, Dimerisation mutants of lac repressor. I. A monomeric mutant, L251A, that binds Lac operator DNA as a dimer, J. Mol. Biol. 290 (1999) 653–666. http://dx.doi.org/10.1006/jmbi.1999.2902.

[76] A. Schmitz, U. Schmeissner, J.H. Miller, Mutations affecting the quaternary structure of the lac repressor, J. Biol. Chem. 251 (1976) 3359–3366.

[77] T.J. Daly, K.S. Matthews, Characterization and modification of a monomeric form of the lactose repressor protein, Biochemistry 25 (1986) 5474–5478.

[78] C.M. Falcon, L. Swint-Kruse, K.S. Matthews, Designed disulfide between N-terminal domains of lactose repressor disrupts allosteric linkage, J. Biol. Chem. 272 (1997) 18550–18557.

[79] S. Semsey, M.Y. Tolstorukov, K. Virmik, V.B. Zhurkin, S. Adhya, DNA trajectory in the Gal repressosome, Genes Dev. 13 (1999) 1251–1262.

[80] F.L. Sousa et al. / Data in Brief 8 (2016) 948–957.
[93] V. Chaptal, V. Gueguen-Chaignon, S. Poncet, C. Lecampion, P. Meyer, J. Deutscher, et al., Structural analysis of Bacillus subtilis CcpA effector binding site, Proteins 64 (2006) 814–816. http://dx.doi.org/10.1002/prot.21001.

[94] M.A. Schumacher, G. Seidel, W. Hillen, R.G. Brennan, Phosphoprotein Crh-Ser46-P displays altered binding to CcpA to effect carbon catabolite regulation, J. Biol. Chem. 281 (2006) 6793–6800.

[95] C.E. Bell, M. Lewis, Crystallographic analysis of lac repressor bound to natural operator O1, J. Mol. Biol. 312 (2001) 921–926.

[96] C.E. Bell, M. Lewis, A closer view of the conformation of the lac repressor bound to operator, Nat. Struct. Biol. 7 (2000) 209–214.

[97] U. Hars, R. Horlacher, W. Boos, W. Welte, K. Diederichs, Crystal structure of the effector-binding domain of the trehalose-repressor of E. coli, a member of the LacI family, in its complexes with inducer trehalose-6-phosphate and noninducer trehalose, Protein Sci. 7 (1998) 2511–2521.

[98] M.A. Schumacher, A. Glasfeld, H. Zalkin, R.G. Brennan, The X-ray structure of the PurR-guanine-purF operator complex reveals the contributions of complementary electrostatic surfaces and a water-mediated hydrogen bond to corepressor specificity and binding affinity, J. Biol. Chem. 272 (1997) 22648–22653.

[99] M. Lewis, G. Chang, N.C. Horton, M.A. Kercher, H.C. Pace, M.A. Schumacher, et al., Crystal structure of the lactose operon repressor and its complexes with DNA and inducer, Science 271 (1996) 1247–1254.

[100] M.A. Schumacher, K.Y. Choi, F. Lu, H. Zalkin, R.G. Brennan, Mechanism of corepressor-mediated specific DNA binding by the purine repressor, Cell 83 (1995) 147–155.

[101] A.M. Friedman, T.O. Fischmann, T.A. Steitz, Crystal structure of lac repressor core tetramer and its implications for DNA looping, Science 268 (1995) 1721–1727.

[102] M.A. Schumacher, K.Y. Choi, H. Zalkin, R.G. Brennan, Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices, Science 266 (1994) 763–770.

[103] H.M. Berman, J. Westbrook, Z. Feng, G. Gililand, T.N. Bhat, H. Weissig, et al., The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.