


# The C repressor of the P2 bacteriophage

Tariq Massad<sup>1</sup> · Evangelos Papadopoulos<sup>1,2</sup> · Pål Stenmark<sup>1</sup> · Peter Damberg<sup>3,4</sup> 

Received: 18 August 2015 / Accepted: 26 October 2015 / Published online: 22 December 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

## Biological context

The C-repressor of bacteriophage P2 (P2 C) is a DNA-binding protein that controls the lifecycle of the P2 bacteriophage by directing it toward the lysogenic mode. P2 C is a 99 amino acids protein, which forms stable homodimers but not higher oligomers in the absence of DNA (Ahlgren-Berg et al. 2007). As opposed to the more common situation where dimeric proteins bind palindromic DNA-sequences, P2 C binds cooperatively to two direct repeats of DNA (named O1 and O2) flanking the –10 region in the Pe promoter in the genetic switch of the bacteriophage P2. The two 8 basepairs long direct repeats have a centre-to-centre distance of 22 base pairs (Ljungquist et al. 1984). According to a Electrophoretic Mobility Shift Assay (EMSA) analysis (Ahlgren-Berg et al. 2007), P2 C induces a high degree of bending of DNA upon binding. The puzzling question how a symmetric protein dimer can bind to an asymmetric DNA binding site where the epitope is repeated twice, as opposed to the more common inverted repeats. To the best of the authors' knowledge, there are only three other examples of proteins

that bind direct repeated DNA sequences in the protein databank. Those are the  $\lambda$ -CII (Jain et al. 2005), the  $\omega$ -repressors (Weihofen et al. 2006) and the mammalian HOT1 (Kappei et al. 2013), which are all structural different to P2C (Massad et al. 2010). The DNA-binding epitope of P2 C is located in the N-terminus (residues 1–54), which contains a helix-turn-helix (HTH) motif (Eriksson et al. 2000; Massad et al. 2010). It has been reported that upon the superinfection of the satellite bacteriophage P4 of a P2 lysogenic cell, P4 is able to derepress the P2 lysogen (Liu and Haggård-Ljungquist 1999). This is mediated by binding the P4 E antirepressor to the P2 C after infection leading to the formation of multimeric complexes, thereby preventing the P2 C from binding to its operator (Liu and Haggård-Ljungquist 1999).

Several mutations have been done on P2 C combined with activity assays to study the C-termini, the dimerization interface and the HTH motif, and to study the deactivation of P2 C by the P4 E antirepressor (Eriksson et al. 2000; Massad et al. 2010). One of the most interesting mutations is the truncation mutation performed on the last 9 residues of the C-terminus, which proved that the P2 C is still active even after truncation, indicating that the C-terminus might not be directly involved in the interaction with DNA. Solving the 3D structure of P2 C improves our understanding of its function and it is the first step to determine its DNA-binding mode.

The backbone assignment of the P2 C has been published and deposited in the Biological Magnetic Resonance Bank (BMRB) under accession code 15577 (Massad et al. 2008). Here we report the solution structure of the P2 C together with the order parameters calculated from <sup>15</sup>N relaxation data using the model-free approach. We have previously reported the crystal structure (PDB 2XCJ) of P2 C at 1.8 Å (Massad et al. 2010), where P2 C was shown to

✉ Peter Damberg  
peter.damberg@ki.se

<sup>1</sup> Department of Biochemistry and Biophysics, Stockholm University, 10691 Stockholm, Sweden

<sup>2</sup> Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Av, Boston, MA 02115, USA

<sup>3</sup> Department of Clinical Neurosciences, Karolinska Institute, 171 77 Stockholm, Sweden

<sup>4</sup> Karolinska Research and Imaging Center, Karolinska University Hospital, 171 76 Stockholm, Sweden

be in a homodimeric state. The crystal structure indicated five rigid helices in the N-terminus and a  $\beta$ -turn in the C-terminus. Since P2 C is a homodimeric protein in the absence of DNA, its dimer interface in solution has been determined with aid from the crystal structure assuming no conformational changes during the crystallization process.

## Methods and results

An *E. coli* strain BL21(DE3) containing plasmid pEE679 expressing P2 C was grown at 310 K in M9 minimal medium containing  $^{13}\text{C}$  labeled-glucose,  $^{15}\text{N}$  labeled- $\text{NH}_4\text{Cl}$  and ampicillin (100 mg/ml) for 6–8 h until an  $\text{OD}_{600} = 0.6$  was reached. Protein expression was induced by addition of isopropyl  $\beta$ -D-thiogalactoside (IPTG) to a final concentration of 1 mM at 37 °C for 4 h. The cells were harvested by centrifugation for 20 min at 9,000g at 4 °C and resuspended in 10 mM sodium phosphate buffer, pH 7.0. Cells were lysed by freezing/thawing together with sonication and thereafter centrifuged at 31,000g for 15 min at 277 K. The supernatant was collected and filtered with a 0.45  $\mu\text{m}$  filter before starting the purification process. The protein was purified using ÄKTA™ FPLC-system in three consecutive steps. First, the filtered sample was adjusted to pH 8.0 with 5 M NaOH and loaded on a weak anion exchange column (DEAE, GE Healthcare) that had been equilibrated with 10 mM sodium phosphate buffer, pH 7.0 (running buffer). P2 C elutes with the flow through, as the pH of the running buffer is lower than the pI of P2 C. The second step was affinity chromatography using a HiTrap Heparin HP column equilibrated with running buffer. P2 C was eluted by a nine-column volume gradient of 1 M NaCl. The eluted fractions contain P2 C were loaded on a Superdex 200 gel filtration column (GE Healthcare) for further purification using 10 mM Na-Phosphate buffer, pH 7.0, 150 mM NaCl as running buffer. Finally, the sample was concentrated to 6 mg/ml using Amicon Ultra-15 centrifugal tubes (Millipore) with molecular weight cutoff 5 kDa.  $\text{D}_2\text{O}$  was added to a final concentration of 10 % before the protein was transferred into a 5 mm NMR tube.

In order to obtain the NOE constraints, the 3D  $^{13}\text{C}$ -NOESY-HSQC and  $^{15}\text{N}$ -NOESY-HSQC (Zhang et al. 1994) datasets with 150 ms mixing time each were acquired at 310 K using Varian INOVA 900 MHz at the Swedish NMR Centre, Göteborg, Sweden. Standard  $^{13}\text{C}$  filtered NOESY experiments were performed to determine the dimer interface where  $^{15}\text{N}$ ,  $^{13}\text{C}$  labeled P2 C was mixed with a nonlabeled variant. Those experiments failed possibly due to insensitivity of the experiments, the size of the protein or unfavorable exchange times, i.e. if the final concentration of mixed labeled:unlabeled dimers was too low. Spectra were processed using NMRPipe (Delaglio

et al. 1995) and analyzed using Sparky (Goddard and Kneller). Model structures were generated using CYANA2.1 software (Güntert 2004). TALOS (Cornilescu et al. 1999) was utilized to generate empirical dihedral angles constraints for 75 residues based on the sequence and chemical shift assignment. A total of 1791 NOEs were manually assigned. The dimer interface of the crystal structure was examined for potential inter domain NOEs. The interdomain pairs M66-methyl-M50H $^{\gamma}$ , L63H $^{\gamma}$ -T62-methyl, A71methyl-S74H $^{\beta}$ , I5methyl-D47-H $^{\beta}$  & T67H $^{\gamma}$ -T59H $^{\gamma 2}$  were identified as connectivities, which would display intense interdomain NOEs but not intradomain NOEs based on the distance. Out of the 1791 manually assigned NOEs those five were specified as interdomain, while the other were treated as ambiguous in the first CYANA iteration. A total of 100 structures were generated for each of the seven CYANA iterations and best structures were selected for the next iteration. Through network anchoring algorithm CYANA identified 3887 intradomain NOEs, as well as 128 interdomain NOEs, where the latter were all manually verified by examining the spectra to ensure reliability. Table 1 shows the structural statistics of P2 C.

The solution structure of P2 C reveals five short  $\alpha$ -helices in each monomeric unit; I5-E16 (helix 1), R20-T26 (helix 2), Y31-S39 (helix 3), T46-Q54 (helix 4), Q57-M66 (helix 5) and a  $\beta$ -sheet-like structure made up by residues

**Table 1** Structural Statistics for P2 C

Average target function ( $\text{Å}^2$ )	2.9 $\pm$ 0.18
<i>Upper distance limits</i>	
Total	4015
Short-range, $ i - j  \leq 1$	1653
Medium-range, $1 <  i - j  < 5$	1043
Long-range, $ i - j  \geq 5$	1191
Interdomain	128
<i>Average RMSD to mean</i>	
Backbone	0.34 $\pm$ 0.12
Heavy atoms	0.78 $\pm$ 0.1
<i>PROCHECK statistics</i>	
Number of conformers	20
Restraints per residue	24.6
RMS deviation for bond lengths	0.001 $\text{Å}$
RMS deviation for bond angles	0
Average of bad steric contacts/100 residues	0
<i>Ramachandran quality (1–81)</i>	
Residues in most-favored regions	85 %
Residues in additionally allowed regions	15 %
Residues in generously allowed regions	0 %
Residues in disallowed regions	0.1 %

Q69–Q76. The C-terminus (starting from residue H85) is shown to be fully flexible and unstructured from the chemical shift index and  $^{15}\text{N}$ -relaxation data. Figure 1 shows stereo views of the ensemble of structures with lowest target function.

### Backbone dynamics of P2 C

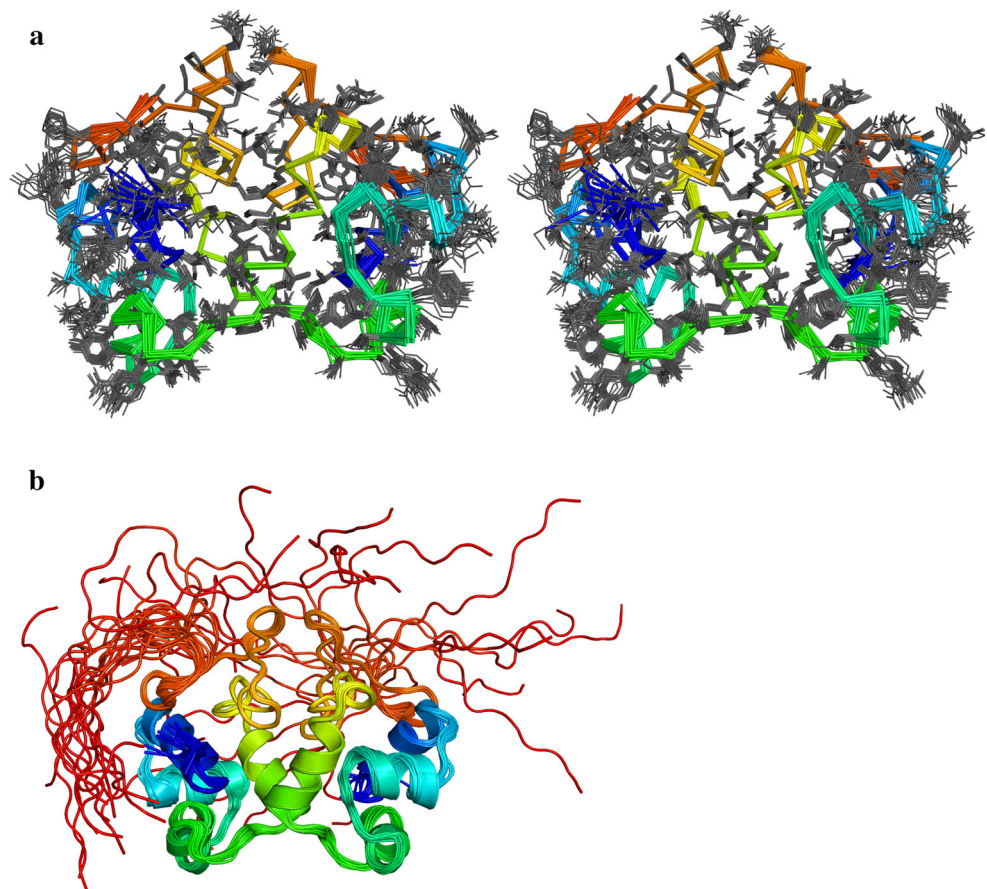
The internal flexibility of P2 C was investigated by measuring  $^{15}\text{N}$  T1,  $^{15}\text{N}$  T2 and  $^{15}\text{N}$ -H NOE relaxation data at a 700 MHz magnetic field (CREM, Florence, Italy). The model-free formalism was used to analyze the dynamics assuming isotropical tumbling with  $\Delta\sigma = 169$ ,  $r_{\text{H-N}} = 1.02$  Å (Damberg et al. 2005) and an overall correlation time of 13.5 ns, as calculated for a 22 kDa globular protein was used in the analysis. The model-free fits were generated using FastmodelFree software (Cole and Loria 2003). The generalized order parameter  $S^2$  is successfully calculated for 69 residues that show intense well-resolved  $^{15}\text{N}$ -HSQC peaks. The  $^{15}\text{N}$  relaxation data shows that the core of the protein (residues 5–85) is rigid with a mean  $S^2 = 0.85$ . The C-terminus (residues 87–99) of the protein has much lower  $S^2$  values ranging from 0.26 to 0.048

which is an indication of considerable flexibility, while the short turns connecting the  $\alpha$ -helices and the  $\beta$ -sheet like structure are rigid on the ns-ps timescale (Fig. 2).

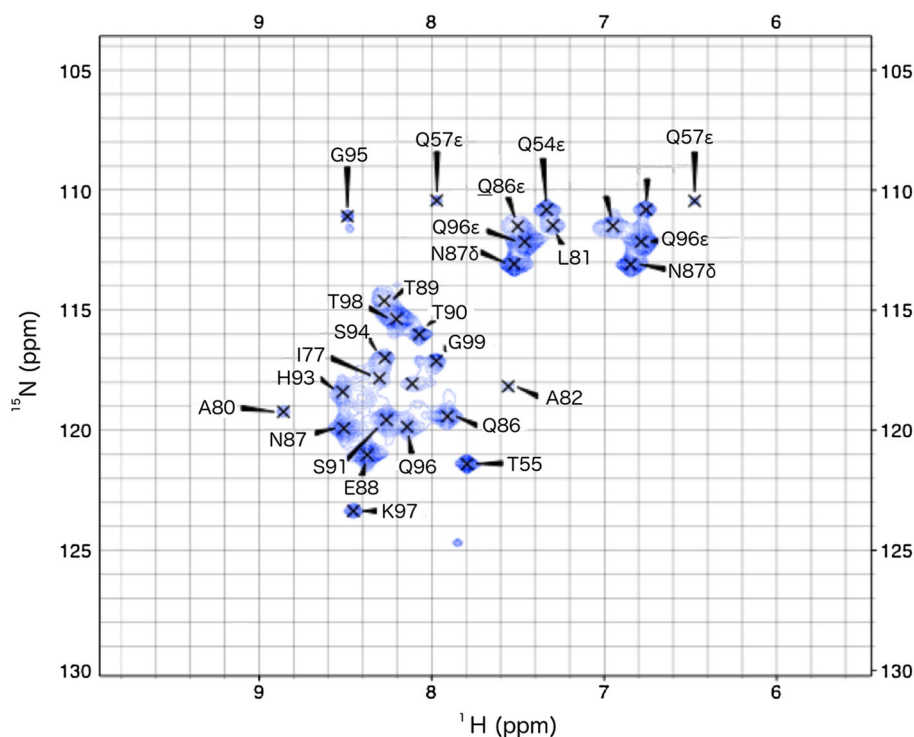
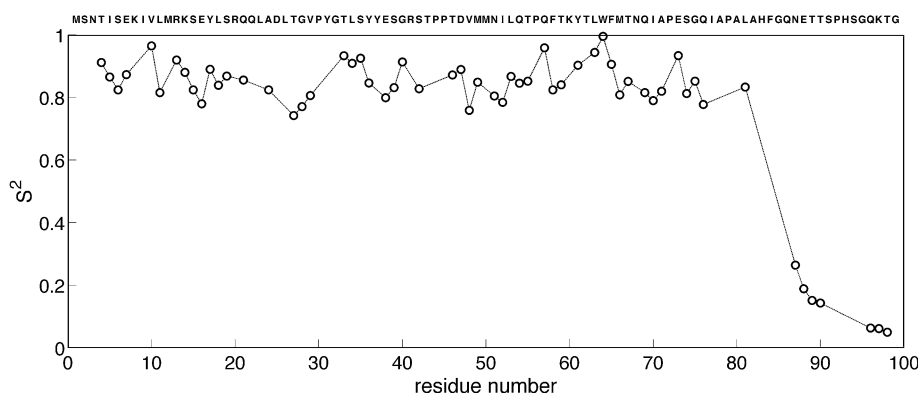
### P2 C-DNA complex

Wildtype  $^{15}\text{N}$ -labeled P2 C in NMR buffer (10 mM Sodium Phosphate, pH = 6.0) was mixed with a 42 bp piece of double stranded DNA with the sequence CATGGTGT**TTAGAT**CTCAATAGTATTTAG**TTTAG**ATGTAGAT and the complementary strand (Sigma) contains both O1 and O2 half sites (underlined and bold) at 278 K for 60 min. The molar ratio between P2 C and the DNA was 4 monomers: 1 stretch of DNA. The mixture was injected into a 3 mm NMR tube and  $\text{D}_2\text{O}$  was added to a final concentration of 10 % and the concentration of double stranded DNA and monomeric protein was 50 and 200  $\mu\text{M}$ , respectively.  $^{15}\text{N}$ -HSQC was recorded on an 800 MHz Bruker AVANCE spectrometer at 298 K, equipped with a *cryo* probe. The spectrum was processed with Topspin 2.1 (Bruker). The  $^{15}\text{N}$ -HSQC spectrum of the complex (Fig. 3) contained 29 peaks, which are tentatively assigned based on

**Fig. 1** Rainbow colored overlays of the backbone of the 20 conformations of lowest target function with blue N-termini and red C-termini. **a** Stereo views of Ca-trace and heavy atoms from the side-chains of residues 1–85 **b** ribbon representation illustrating the flexible C-termini



**Fig. 2** The generalized order parameter  $S^2$  of P2 C. The average  $S^2$  value for the C-terminus (residues 5–85) is 0.85 corresponding to rigid segments. The N-terminus displays values corresponding to highly flexible residues



**Fig. 3**  $^{15}\text{N}$ -HSQC of P2 C DNA Complex. Only peaks corresponding to the flexible c-terminus are observed in the spectrum

similarity of chemical shifts. The most C-terminal residues (85–99) display signals.

## Discussion and conclusions

We have determined the solution structure of the dimeric P2 C repressor protein. The solution structure of P2 C displays close agreement with the crystal structure with a backbone RMSD of 1.16 Å between the crystal structure and the NMR-structure with the lowest CYANA target function for residues 5–81. This is crucial for the structure calculations since the five manually assigned inter domain NOEs were inferred based on shorter inter atomic distances for inter domain pairs. If the crystal and solution structures were not similar

those five assignments would potentially be incorrect. Incorrect inter domain constraints would guide the structure calculation towards an incorrect local minimum of the target function. In addition to the a priori argument for similarity between the crystal and solution structures, such as secondary chemical shifts,  $^{15}\text{N}$ -relaxation data and mutation studies, a large body of evidence accumulates during the structure calculation. In particular more than 4000 NOEs including more than 100 inter domain and the convergence to a tight ensemble of structures strikingly similar to the crystal structure support the initial assumption.

The final 20 lowest target function structures were evaluated using PROCHECK\_NMR (Morris et al. 1992). All dihedral angles for structured residues are in allowed conformations of Ramachandran map (Lovell et al. 2003

The C-terminus is shown to be flexible in solution from NMR data with very low order parameters, random coil chemical shift index and the absence of NOESY peaks. In the crystal structure no electron density is observed for the C-terminal residues after G85, also indicating that the C-terminal is disordered.

The C-terminus appears flexible, also in the  $^{15}\text{N}$  relaxation analysis, while the well-folded part of the sequence (residues 4–81) appears rigid. For the flexible C-terminus (relaxation data from residues 87–98) the extended model free model (Clare et al. 1990), i.e. model 4, is preferred. The internal correlation times are several hundred picoseconds and generalized squared order parameters for the slow internal component, i.e.  $S_s^2$  are below 0.33 and the fast components,  $S_f^2$ , are in the range of 0.71–0.81. For the well-folded part the F-test indicates that model 1 is preferred with  $S^2$  in the range 0.74–0.96. The turn connecting helices 2 and 3 displays lower  $S^2$ , while other turns appear as rigid as the elements of secondary structure on the picoseconds to nanosecond timescale. For residues T4 and F5 in the N-terminus, the model-free model (Lipari and Szabo 1982), i.e. model 2, is preferred with order parameters of 0.83 or higher and internal correlation times in the tens of picoseconds regime. For four residues (N51, I52, F65 and M66) the F-test indicates that significant exchange broadening contributes to R2 (the exchange contributions to R2 are 11, 13, 20 and 16  $\text{s}^{-1}$  for N51, I52, F65 and M66, respectively). It is noteworthy that the local structures of the residues displaying exchange broadening are somewhat different in the crystal and solution structures. XTLSSR (King and Johnson, 1999) identifies N51 and I52 as members of a  $3_{10}$ -helix in 23 % of the members of the ensemble, while they are classified as  $\alpha$ -helical in the other ensemble members. In the crystal structure (2xcj) XTLSSR classifies them as  $\alpha$ -helical. Residues F65 and M66 are classified as hydrogen bonded turn in most ensemble members, while they are classified as  $\alpha$ -helical in 18 % of the members. In the crystal structure they are classified as  $\alpha$ -helical. It is tempting to hypothesize that the observed uncertainties in the local structure for those sites are genuine features of the protein as they are consistent with exchange broadening.

The observable HSQC-peaks in the complex with DNA demonstrate that the C-terminus remains flexible also in the presence of DNA. This may be somewhat surprising as the C-terminus displays significant sequence identity to C-proteins from related phages (Massad et al. 2010). However, the finding that the C-terminus is flexible also in the complex with DNA explain an in vitro activity assays which demonstrated that a C-terminally truncated variant P2 C (1–90) is capable of binding to the target DNA and function as a repressor of a reporter gene (Massad et al. 2010).

A few signals were tentatively assigned to residues part of the helix bundle (T55, L81 and A82). If correctly assigned, this would indicate a significantly increased flexibility of those residues upon binding to DNA. The absence of HSQC peaks from the rigid core of the protein is caused by rapid relaxation, likely caused by slow tumbling of the 66 kDa complex and hence provides some evidence in support of the tetrameric binding model. However, exchange between complexes with different stoichiometry could lead to signal loss through exchange broadening, and cannot be completely ruled out based on the current data. Solving the crystal and solution structures of P2 C has opened way for many questions regarding how a small protein like P2 C can bind such long DNA stretch (the center-to-center distance between O1 and O2 half sites is 22 bp). In addition, having the C-terminus very flexible even upon binding the DNA raises questions regarding its biological role.

**Acknowledgments** We acknowledge Professor Elisabeth Haggård-Ljungquist for the important discussions and experimental assistance in the P2 C overexpression and purification. Dr. Isabella Felli and the CERM Magnetic Resonance Center in Florence is gratefully acknowledged for kind assistance and use of their NMR equipment. Professor Göran Karlsson from the Swedish NMR centre is acknowledged for the assistance in the NOESY experiments. Carl Trygger and Magn. Bergvall foundations are acknowledged for financial support to P.D. This work was also supported by grants from the Swedish Research Council (2010-5200 and 2014-5667), the Swedish cancer society, the Carl Trygger, Magn. Bergvall and Wenner-Gren foundations and to P.S.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Ahlgren-Berg A, Henriksson-Peltola P, Sehlen W, Haggård-Ljungquist E (2007) A comparison of the DNA binding and bending capacities and the oligomeric states of the immunity repressors of heteroimmune coliphages P2 and WPhi. *Nucleic Acids Res* 35:3167–3180. doi:10.1093/nar/gkm171
- Clare GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J Am Chem Soc* 112:4989–4991
- Cole R, Loria JP (2003) FAST-modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. *J Biomol NMR* 26:203–213
- Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289–302
- Damberg P, Jarvet J, Gräslund A (2005) Limited variations in  $^{15}\text{N}$  CSA magnitudes and orientations in ubiquitin are revealed by

- joint analysis of longitudinal and transverse NMR relaxation. *J Am Chem Soc* 127:1995–2005
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293
- Eriksson SK, Liu T, Haggård-Ljungquist E (2000) Interacting interfaces of the P4 antirepressor E and the P2 immunity repressor C. *Mol Microbiol* 36:1148–1155
- Goddard TD, Kneller GD SPARKY 3. University of California, San Francisco
- Güntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* 278:353–378. doi:10.1385/1-59259-809-9:353
- Jain D, Kim Y, Maxwell KL, Beasley S, Zhang R, Gussin GN, Edwards AM, Darst SA (2005) Crystal structure of bacteriophage lambda cII and its DNA complex. *Mol Cell* 19:259–269. doi:10.1016/j.molcel.2005.06.006
- Kappei D, Butter F, Benda C, Scheibe M, Draskovic I, Stevense M, Novo CL, Basquin C, Araki M, Araki K, Blazhev D, Kittler R, Jessberger R, Londoño-Vallejo JA, Mann M, Buchholz F (2013) HOT1 is a mammalian direct telomere repeat-binding protein contributing to telomerase recruitment. *EMBO J* 32:1681–1701
- King SM, Johnson WC (1999) Assigning secondary structure from protein coordinate data. *Proteins* 3(35):313–320
- Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* 104:4546–4559
- Liu T, Haggård-Ljungquist E (1999) The transcriptional switch of bacteriophage WPhi, a P2-related but heteroimmune coliphage. *J Virol* 73:9816–9826
- Ljungquist E, Kockum K, Bertani LE (1984) DNA sequences of the repressor gene and operator region of bacteriophage P2. *Proc Natl Acad Sci USA* 81:3988–3992
- Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by calpha geometry: Phi, psi and cbeta deviation. *Proteins* 50:437–450. doi:10.1002/prot.10286
- Mandel AM, Akke M, Palmer AG III (1995) Backbone dynamics of escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol* 246:144–163
- Massad T, Papadopoulos E, Henriksson-Peltola P, Haggård-Ljungquist E, Damberg P (2008) Assignment of <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N chemical shift resonances of P2 C-repressor protein. *Biomol NMR Assign* 2:215–217. doi:10.1007/s12104-008-9124-6
- Massad T, Skaar K, Nilsson H, Damberg P, Henriksson-Peltola P, Haggård-Ljungquist E, Hogbom M, Stenmark P (2010) Crystal structure of the P2 C-repressor: a binder of non-palindromic direct DNA repeats. *Nucleic Acids Res*. doi:10.1093/nar/gkq626
- Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. *Proteins* 12:345–364. doi:10.1002/prot.340120407
- Weihofen WA, Cicek A, Pratto F, Alonso JC, Saenger W (2006) Structures of omega repressors bound to direct and inverted DNA repeats explain modulation of transcription. *Nucleic Acids Res* 34:1450–1458. doi:10.1093/nar/gkl015
- Zhang O, Kay LE, Olivier JP, Forman-Kay JD (1994) Backbone <sup>1</sup>H and <sup>15</sup>N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J Biomol NMR* 4:845–858