DNA Crystallography

DOI: 10.1002/anie.201409014

Racemic DNA Crystallography**

*Pradeep K. Mandal, Gavin W. Collie, Brice Kauffmann, and Ivan Huc**

Abstract: *Racemates increase the chances of crystallization by allowing molecular contacts to be formed in a greater number of ways. With the advent of protein synthesis, the production of protein racemates and racemic-protein crystallography are now possible. Curiously, racemic DNA crystallography had not been investigated despite the commercial availability of* L- *and* D-*deoxyribo-oligonucleotides. Here, we report a study into racemic DNA crystallography showing the strong propensity of racemic DNA mixtures to form racemic crystals. We describe racemic crystal structures of various DNA sequences and folded conformations, including duplexes, quadruplexes, and a four-way junction, showing that the advantages of racemic crystallography should extend to DNA.*

Racemic mixtures of small organic molecules have a well-known tendency to crystallize as racemic crystals and to much less frequently separate into conglomerates in which each crystal is composed of a single enantiomer.[1] The observation that racemic crystals tend to be denser than their chiral counterparts, known as the Wallach rule, was made as early as 1895.[2] The physical basis of this close-packing phenomenon and whether it is truly related to the higher occurrence of racemates is still debated.[3] Nevertheless, the prevalence of racemates has been repeatedly verified experimentally over the years. It applies to small molecules in which chirality arises from stereogenic centers and also to helical molecules such as helicates[4] and helical aromatic or aliphatic foldamers,[5,6] but not to unsubstituted helicenes, which often form conglomerates.[7] The propensity of enantiomers to cocrystallize is so strong that even pseudo-enantiomers (molecules that are almost but not exactly mirror images) cocrystallize to form so-called quasi-racemic crystals.[8] With the advent of protein chemical synthesis,[9] racemic protein mixtures have been produced and repeatedly shown to cocrystallize as racemic crystals,[10] especially in the $P1$ (or $P\bar{1}$) space group. Theoretical calculations had predicted this tendency based on entropic considerations,[11] as opposed to molecular close-packing invoked in the case of small organic molecules. These properties have been turned into an advantage for protein crystallogenesis, as they offer additional opportunities for crystal growth, in which single enantiomers fail to produce X-ray quality crystals.[10,12] Quasi-racemic protein crystallography has also been taken advantage of.[13]

Curiously, it appears that racemic DNA crystallography has not yet been investigated and the potential benefits of this approach to elucidate nucleic acid structures are not known. Only one structure of a racemic DNA duplex has been reported in the protein data bank,[14a] but no comparison with the structure of the pure D-enantiomer was provided. One racemic RNA duplex has also been described.[14b] A possible reason for this apparent discrepancy between protein and DNA racemic crystallography may be that L-oligonucleotides used to be expensive, especially regarding the amounts required for crystallogenesis. However, due to their nuclease resistance, these nonnatural enantiomers provide interesting applications as spiegelmers[15] and spiegelzymes.[16] Consequently, their production has increased in recent years and production costs have dropped to levels not much higher than those of D-DNA. We thus endeavored to explore in a more systematic manner how racemic mixtures of DNA oligonucleotides could crystallize and now report their strong tendency to form well-ordered racemic crystals, suggesting that the advantages of racemic protein crystallography may be successfully applied to other biomolecules.

To investigate the applicability of a racemic crystallization approach to DNA crystallography, a variety of DNA sequences were chosen with the aim of covering a diverse range of sequences and structures. Additionally, sequences were selected for which at least one crystal structure of the natural D-enantiomer was known,[17–19] thereby permitting the validation of any racemic crystal structures arising from our studies, as well as allowing us to compare crystal growth conditions. The sequences selected (see Table 1) encompass deoxyribo-oligonucleotides of 6, 10, and 12 nucleotides in length, the natural D-enantiomers of which fold as tetramolecular and bimolecular quadruplexes, B-type duplexes, and four-way junctions. Deoxyribo-oligonucleotide sequences were purchased from commercial suppliers, with a 3 μmol synthesis scale typically providing sufficient material for crystallographic studies. Racemic solutions of each DNA

[*] Dr. P. K. Mandal, Dr. G. W. Collie, Dr. I. Huc
Université de Bordeaux, CBMN, UMR5248
Institut Européen de Chimie et Biologie
2 rue Robert Escarpit, 33600 Pessac (France)
and
CNRS, CBMN, UMR5248
E-mail: i.huc@iecb.u-bordeaux.fr
Dr. B. Kauffmann
Université de Bordeaux, UMS3033
Institut Européen de Chimie et Biologie (IECB)
2 rue Robert Escarpit, 33600 Pessac (France)
and
CNRS, IECB; UMS3033
and
INSERM, IECB, UMS3033

[**] This work was supported by the European Union's Seventh Framework Program through the European Research Council (grant agreement no. ERC-2012-AdG-320892, postdoctoral fellowship to P.K.M.). We thank Dr. B. Baptiste for preliminary investigations. We thank SOLEIL and the European Synchrotron Radiation Facility for access to synchrotron radiation facilities (beam lines Proxima-1 and ID23-2, respectively), and we thank Dr. Pierre Legrand for assistance on Proxima-1.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201409014.

Table 1: Crystal structures obtained from racemic mixtures of deoxyribo-oligonucleotides.

Sequence[a]	Structural motif	Max. res. [Å][b]	Space group	PDB entry
D/L-TG_4T (K^+)	tetramolecular G-quadruplex	2.69	$P\bar{1}$	4R44
D/L-$G_4T_4G_4$ (K^+) form-1	bimolecular G-quadruplex	1.90	$P\bar{1}$	4R45
D/L-$G_4T_4G_4$ (K^+) form-2	bimolecular G-quadruplex	1.85	$P\,2_1/n$	4R47
D/L-CCGGTACCGG (Ca^{2+})	four-way junction	2.33	$C2/c$	4R48
D/L-CCGGTACCGG (Ca^{2+})	B-type duplex	1.28	$P\bar{1}$	4R49
D/L-CCGGTACCGG (Co^{2+})	B-type duplex	1.49	$P\bar{1}$	4R4A
D/L-CCGGTACCGG (Mg^{2+})	B-type duplex	1.29	$P\bar{1}$	4R4D

[a] Cations in parenthesis are included in the structure. [b] Maximal resolution.

sequence were prepared by mixing equal amounts of pure D- and L-enantiomers. Crystallization trials were then performed for each racemic DNA mixture, using standard hanging-drop vapor diffusion methods. Crystallization conditions similar to those reported for the nonracemic sequences were initially tried for all sequences listed in Table 1, which yielded crystals suitable for X-ray diffraction analysis following standard optimization of the crystallization conditions for all sequences except D/L-TG_4T (see Table S1). For this sequence, we first reproduced the crystallization of D-TG_4T under the reported conditions and confirmed the different behavior of the racemate. Since we were not able to grow crystals using the crystallization conditions reported for the enantiopure sequence, a sparse-matrix crystallization screening approach was adopted for the racemate. This identified a condition suitable for crystal growth composed of 1,6-hexanediol, potassium cacodylate, potassium chloride, and magnesium chloride—somewhat different from the crystallization conditions of the native enantiopure sequence (see Table S1). Interestingly, pure D-TG_4T does not crystallize under these conditions. Eventually, crystals of suitable quality for X-ray diffraction analysis were obtained for all DNA sequences listed in Table 1. Images showing crystal form and morphology are shown in Figure 1, and details of crystallization conditions are provided in Table S1.

Preliminary X-ray diffraction analyses of these crystals were performed on an in-house rotating anode diffractometer. Crucially, analysis of the initial X-ray diffraction measurements for each crystal easily identified a center of inversion, with the crystals belonging to space groups $P\bar{1}$, $C2/c$, or $P2_1/n$ (Table 1), all commonly encountered in crystals of small organic molecules. The apparent prevalence of space group $P\bar{1}$ also mirrors racemic protein crystallization experiments. As pointed out by Wukowitz and Yeates,[11] the high level of degrees of freedom, or dimensionality, of the $P\bar{1}$ space group makes it a common and highly favorable packing arrangement. We also note that the predicted next-best packing arrangements consist of $P2_1/n$ and $C2/c$ space groups which are both encountered in our study as well.

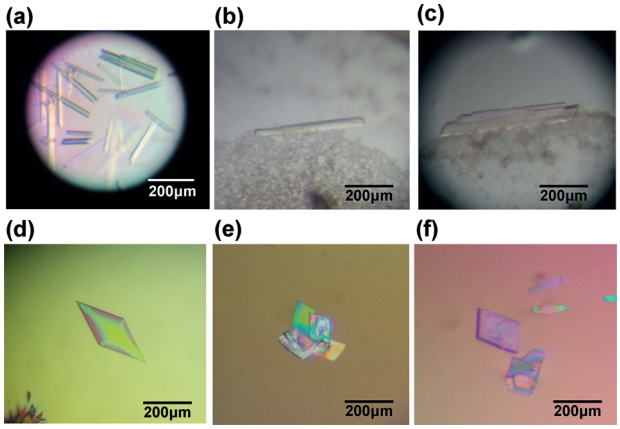

Figure 1. Pictures of crystals obtained for racemic mixtures of DNA sequences: a) D/L-TG_4T; b,c) D/L-$G_4T_4G_4$; d–f) D/L-CCGGTACCGG.

Full datasets were then collected using synchrotron or in-house X-ray sources for crystals of all sequences shown in Table 1. Due to the availability of the equivalent D-enantiomer DNA structures as suitable search models, all of these datasets were successfully solved by molecular replacement. Other phasing methods were not investigated. The crystal structures determined from these racemic DNA crystallization trials yielded a total of seven distinct structural arrangements (see Figure 2), all of which belong to achiral space groups (Tables 1 and S2), that is, space groups that contain mirror planes, glide planes, or inversions as symmetry operations and hence require the presence of both enantiomers of a chiral molecule.[20] Notably, the space groups encountered are all centrosymmetric space groups, a feature that would matter if other phasing methods were to be used. Full details of the X-ray diffraction measurements, structure determination and refinement are described in the Supporting Information. The racemic crystal structures reported here include a tetramolecular G-quadruplex formed from the sequence d(TG_4T), two crystal forms of a bimolecular quadruplex formed from the sequence d($G_4T_4G_4$), plus a four-way junction and three B-type duplex structural arrangements (bound by different divalent cations) all formed from the sequence d(CCGGTACCGG) (Figure 2). In all cases, the D- and L-DNA enantiomers are related by chirality-inverting symmetry operations (i.e., inversions and glide planes) as well as chirality-preserving symmetry operations for space groups $C2/c$ and $P2_1/n$ (i.e., rotations, translations, and screw axes). Importantly, structural alignments of the racemic structures determined in this work compared to the equivalent nonracemic structures reveal only marginal differences (Table S3). This indicates that, despite differences in crystal packing contacts (see below), the three-dimensional structures generated from racemic DNA mixtures have the same relevance and utility as the equivalent D-enantiomer crystal structures.

Inspection of the crystal packing contacts reveals a number of interactions unique to the racemic structures. For example, the racemic $G_4T_4G_4$ G-quadruplex structure (in space group $P\bar{1}$) features an unusual intermolecular thymine–thymine (TT) base pair linking D- and L-molecules in the crystal lattice (see Figure S1). This TT base pair involves the formation of hydrogen bonds between the N3 and O4 groups of D- and L-thymine-8 residues, resulting in a base pair

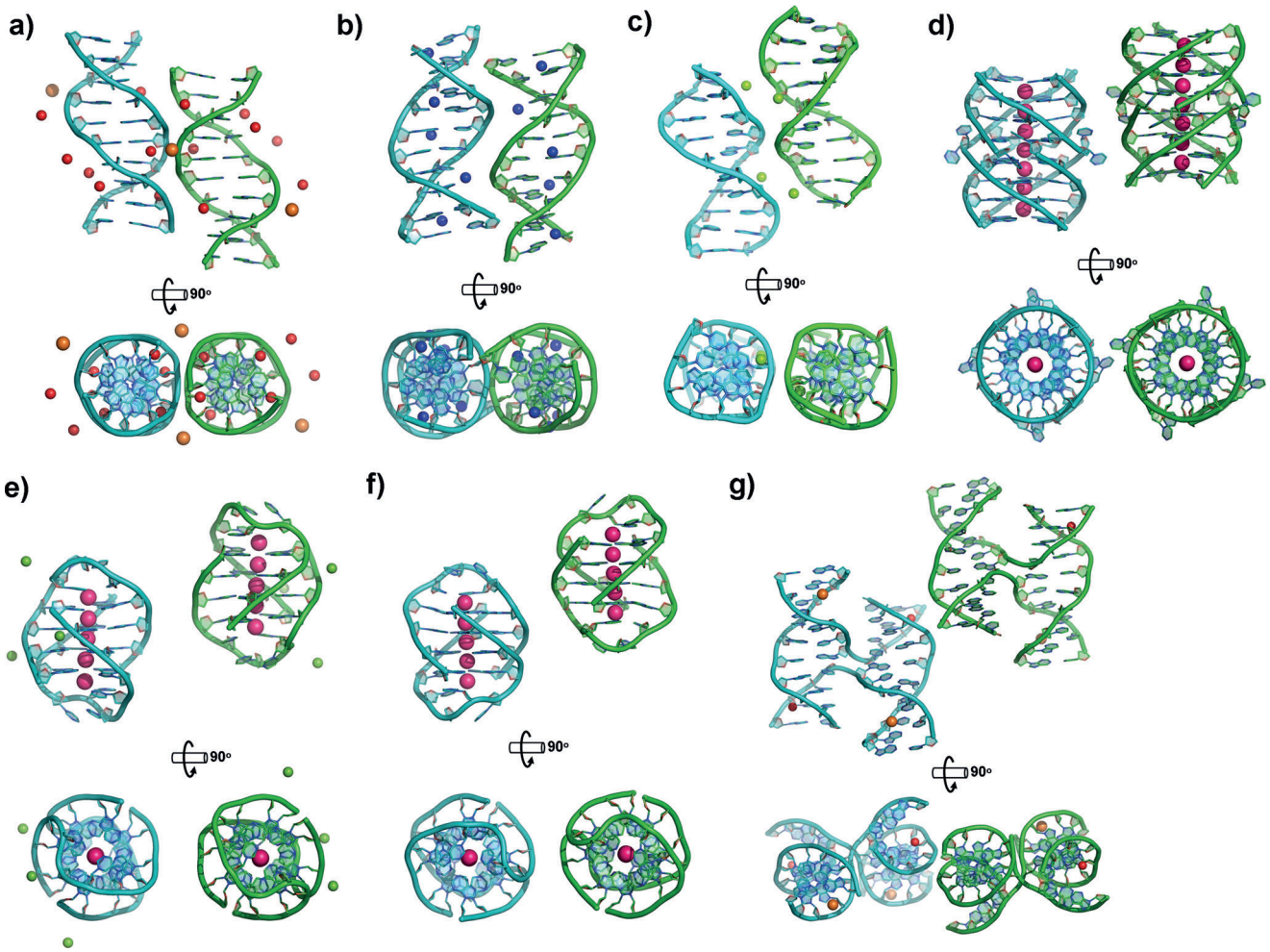

Figure 2. Racemic DNA crystal structures illustrating the relationship between L and D enantiomers (colored cyan and green, respectively). a–c) B-type DNA duplexes formed from the sequence CCGGTACCGG, cocrystallized with either Ca^{2+} (a), Co^{2+} (b), or Mg^{2+} (c) ions. All three duplex structures belong to space group $P\bar{1}$. d) Tetramolecular G-quadruplex formed from the sequence TG_4T in space group $P\bar{1}$. e,f) Bimolecular G-quadruplexes formed from the *Oxytricha nova* telomeric sequence $G_4T_4G_4$ in space groups $P2_1/n$ and $P\bar{1}$, for (e) and (f), respectively. g) Four-way DNA junction formed from the sequence CCGGTACCGG in space group $C2/c$. Ions are represented as spheres, colored orange (sodium), red (calcium), green (magnesium), blue (cobalt), and magenta (potassium).

arrangement comparable to a TT wobble base pair. In comparison, the crystal packing of the natural enantiopure crystal structure (protein data bank, entry 1JRN)[18a] similarly brings two symmetry-related thymine-8 residues within close proximity. However, due to the geometric constraints imposed by the chirality, the thymine bases are unable to form multiple base-pairing hydrogen bonds such as those seen in the racemic structure (Figure S1b). Interestingly, the second racemic crystal form of the bimolecular quadruplex determined in this work (in space group $P2_1/n$) also contains the intermolecular racemic TT base pair seen in the $P\bar{1}$ form, suggesting this achiral motif to be a favorable crystal packing interaction. Further details and discussion of racemic crystal packing contacts in comparison with the corresponding D-DNA crystals are provided in Figures S2–S4. A recurrent trend is a tendency of duplexes, four-way junctions, and TG_4T quadruplexes to form enantiopure pseudohelices by stacking on top of each other. The differences between D- and D/L-DNA crystals then arise from differences in packing one-handed helices only, or a mixture of helices of opposite handedness. In many racemic crystals, enantiopure pseudo-helices of opposite handedness interlock, creating tightly packed lattices (e.g., see Figure S2a).

The ability of opposite DNA enantiomers to form stable intermolecular contacts may contribute to the prevalent occurrence of racemic crystals. Indeed, in no case did we observe the formation of conglomerates (also termed racemic conglomerates),[20] despite the racemic DNA mixtures being exposed to crystallization conditions known to favor crystal growth of the natural enantiopure DNA sequences. To the best of our knowledge, no DNA conglomerate has been reported to date. Racemic crystals are intrinsically different from the pure enantiomer crystals and offer additional and unique packing opportunities unavailable to enantiopure DNA molecules. Thus, racemic DNA mixtures may result in successful crystallogenesis in cases in which enantiopure DNA crystals are unobtainable.

The benefits of applying racemic crystallographic methods to proteins are well known and considerable. We have described here a systematic crystallographic study of a range

of racemic DNA sequences, showing that: a) racemic mixtures of DNA molecules are highly amenable to crystallogenesis; b) these crystals are invariably achiral, showing a preference for the space group $P\bar{1}$, and c) the structures determined from racemic mixtures are nearly identical to those determined from classical enantiopure solutions. These findings indicate that the application of the racemic crystallographic method to DNA structural research is both feasible and potentially highly beneficial. Although beyond the scope of this work, one may predict the future advent of racemic RNA crystallography, and also of quasi-racemic nucleic acid crystallography for the advantages it may bring in terms of phasing.[21]

Received: September 11, 2014
Published online: October 30, 2014

Keywords: crystal growth · DNA · racemic crystallography · structural analysis · X-ray diffraction

[1] For a review, see: a) J. Jacques, A. Collet, S. H. Wilen in *Enantiomers, racemates and resolutions*, 3rd ed., Krieger, Malabar, **1994**.
[2] O. Wallach, *Justus Liebigs Ann. Chem.* **1895**, *286*, 90 – 143.
[3] C. P. Brock, W. B. Schweizer, J. D. Dunitz, *J. Am. Chem. Soc.* **1991**, *113*, 9811 – 9820.
[4] a) J. M. Lehn, A. Rigault, J. Siegel, J. Harrowfield, B. Chevrier, D. Moras, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 2565 – 2569; b) R. Krämer, J. M. Lehn, A. De Cian, J. Fischer, *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 703 – 706; *Angew. Chem.* **1993**, *105*, 764 – 767; c) C. Piguet, G. Bernardinelli, G. Hopfgartner, *Chem. Rev.* **1997**, *97*, 2005 – 2062.
[5] a) Y. Ferrand, A. M. Kendhale, B. Kauffmann, A. Grélard, C. Marie, V. Blot, M. Pipelier, D. Dubreuil, I. Huc, *J. Am. Chem. Soc.* **2010**, *132*, 7858 – 7859; b) M. Kudo, V. Maurizot, B. Kauffmann, A. Tanatani, I. Huc, *J. Am. Chem. Soc.* **2013**, *135*, 9628; c) Q. Gan, Y. Ferrand, N. Chandramouli, B. Kauffmann, C. Aube, D. Dubreuil, I. Huc, *J. Am. Chem. Soc.* **2012**, *134*, 15656; d) G. Lautrette, C. Aube, Y. Ferrand, M. Pipelier, V. Blot, C. Thobie, B. Kauffmann, D. Dubreuil, I. Huc, *Chem. Eur. J.* **2014**, *20*, 1547 – 1553; e) M. Kudo, V. Maurizot, H. Masu, A. Tanatani, I. Huc, *Chem. Commun.* **2014**, *50*, 10090 – 10093; f) M. Lee, J. Shim, P. Kang, I. A. Guzei, S. H. Choi, *Angew. Chem. Int. Ed.* **2013**, *52*, 12564 – 12567; *Angew. Chem.* **2013**, *125*, 12796 – 12799; g) H. Rasmussen, J. S. Kastrup, J. N. Nielsen, J. M. Nielsen, P. E. Nielsen, *Nat. Struct. Biol.* **1997**, *4*, 98 – 101.
[6] For exceptions, see: a) E. Ohta, H. Sato, S. Ando, A. Kosaka, T. Fukushima, D. Hashizume, M. Yamasaki, K. Hasegawa, A. Muraoka, H. Ushiyama, K. Yamashita, T. Aida, *Nat. Chem.* **2011**, *3*, 68 – 73; b) H. Zhao, W. Qiang Ong, F. Zhou, X. Fang, X. Chen, S. F. Y. Li, H. Su, N.-J. Cho, H. Zeng, *Chem. Sci.* **2012**, *3*, 2042 – 2046.
[7] a) W. H. Laarhoven, W. J. C. Prinsen, *Top. Curr. Chem.* **1984**, *125*, 63 – 130; b) R. H. Martin, *Angew. Chem. Int. Ed. Engl.* **1974**, *13*, 649 – 660; *Angew. Chem.* **1974**, *86*, 727 – 738; c) R. H. Martin, M. J. Marchant, *Tetrahedron* **1974**, *30*, 343 – 345; For examples of racemic helicenes, see: d) F. Torricelli, J. Bosson, C. Besnard, M. Chekini, T. Burgi, J. Lacour, *Angew. Chem. Int. Ed.* **2013**, *52*, 1796 – 1800; *Angew. Chem.* **2013**, *125*, 1840 – 1844; e) R. Kuroda, *J. Chem. Soc. Perkin Trans. 2* **1982**, 789 – 794; For bidimensional helicene crystallization, see: f) M. Stöhr, S. Boz, M. Schär, M.-T. Nguyen, C. A. Pignedoli, D. Passerone, W. B. Schweizer, C. Thilgen, T. A. Jung, F. Diederich, *Angew. Chem. Int. Ed.* **2011**, *50*, 9982 – 9986; *Angew. Chem.* **2011**, *123*, 10158 – 10162.
[8] a) H. Jiang, C. Dolain, J.-M. Léger, H. Gornitzka, I. Huc, *J. Am. Chem. Soc.* **2004**, *126*, 1034 – 1035; b) C. Dolain, H. Jiang, J.-M. Léger, P. Guionneau, I. Huc, *J. Am. Chem. Soc.* **2005**, *127*, 12943 – 12951; c) G. Lautrette, B. Kauffmann, Y. Ferrand, C. Aube, N. Chandramouli, D. Dubreuil, I. Huc, *Angew. Chem. Int. Ed.* **2013**, *52*, 11517 – 11520; *Angew. Chem.* **2013**, *125*, 11731 – 11734; d) K. A. Wheeler, R. C. Grove, R. E. Davis, W. S. Kassel, *Angew. Chem. Int. Ed.* **2008**, *47*, 78 – 81; *Angew. Chem.* **2008**, *120*, 84 – 87; e) I. L. Karle, J. Karle, *J. Am. Chem. Soc.* **1966**, *88*, 24 – 27; f) L. Pasteur, *Ann. Chim. Phys.* **1853**, *28*, 437 – 483.
[9] a) P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776 – 779; b) P. E. Dawson, S. B. H. Kent, *Annu. Rev. Biochem.* **2000**, *69*, 923 – 960; c) S. B. H. Kent, *Chem. Soc. Rev.* **2009**, *38*, 338 – 351.
[10] T. O. Yeates, S. B. H. Kent, *Annu. Rev. Biophys.* **2012**, *41*, 41 – 61.
[11] S. W. Wukovitz, T. O. Yeates, *Nat. Struct. Biol.* **1995**, *2*, 1062 – 1067.
[12] a) L. E. Zawadzke, J. M. Berg, *Proteins Struct. Funct. Genet.* **1993**, *16*, 301 – 305; b) C. Toniolo, C. Peggion, M. Crisma, F. Formaggio, X. Shui, D. S. Eggleston, *Nat. Struct. Biol.* **1994**, *1*, 908 – 914; c) B. L. Pentelute, Z. P. Gates, V. Tereshko, J. L. Dashnau, J. M. Vanderkooi, A. Kossiakoff, S. B. H. Kent, *J. Am. Chem. Soc.* **2008**, *130*, 9695 – 9701.
[13] a) R. Okamoto, K. Mandal, M. R. Sawaya, Y. Kajihara, T. O. Yeates, S. B. H. Kent, *Angew. Chem. Int. Ed.* **2014**, *53*, 5194 – 5198; *Angew. Chem.* **2014**, *126*, 5294 – 5298; b) D. E. Mortenson, K. A. Satyshur, I. A. Guzei, K. T. Forest, S. H. Gellman, *J. Am. Chem. Soc.* **2012**, *134*, 2473 – 2476; c) K. Mandal, B. L. Pentelute, D. Bang, Z. P. Gates, V. Y. Torbeev, S. B. H. Kent, *Angew. Chem. Int. Ed.* **2012**, *51*, 1481 – 1486; *Angew. Chem.* **2012**, *124*, 1510 – 1515.
[14] a) M. Doi, M. Inoue, K. Tomoo, T. Ishida, T. Ueda, M. Akagi, H. Urata, *J. Am. Chem. Soc.* **1993**, *115*, 10432 – 10433; b) W. Rypniewski, M. Vallazza, M. Perbandt, S. Klussman, L. J. DeLucas, C. Betzel, V. A. Erdmann, *Acta Crystallogr. Sect. D* **2006**, *62*, 659 – 664.
[15] S. Leva, A. Lichte, J. Burmeister, P. Muhn, B. Jahnke, D. Fesser, J. Erfurth, P. Burgstaller, S. Klussmann, *Chem. Biol.* **2002**, *9*, 351 – 359.
[16] a) B. Wlotzka, S. Leva, B. Eschgfaller, J. Burmeister, F. Kleinjung, C. Kaduk, P. Muhn, H. Hess-Stump, S. Klussmann, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8898 – 8902; b) E. Wyszko, M. Szymanski, H. Zeichhardt, F. Muller, J. Barciszewski, V. A. Erdmann, *Plos One* **2013**, *8*, e54741.
[17] a) K. Phillips, Z. Dauter, A. I. Murchie, D. M. Lilley, B. Luisi, *J. Mol. Biol.* **1997**, *273*, 171 – 182; b) C. Caceres, G. Wright, C. Gouyette, G. N. Parkinson, J. A. Subirana, *Nucleic Acids Res.* **2004**, *32*, 1097 – 1102; c) M. P. Lee, G. N. Parkinson, P. Hazel, S. Neidle, *J. Am. Chem. Soc.* **2007**, *129*, 10106 – 10107.
[18] a) S. Haider, G. N. Parkinson, S. Neidle, *J. Mol. Biol.* **2002**, *320*, 189 – 200; b) P. Hazel, G. N. Parkinson, S. Neidle, *J. Am. Chem. Soc.* **2006**, *128*, 5480 – 5487.
[19] a) B. F. Eichman, J. M. Vargason, B. H. Mooers, P. S. Ho, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3971 – 3976; b) S. Venkadesh, P. K. Mandal, N. Gautham, *Nucleos. Nucleot. Nucl. Acids* **2012**, *31*, 184 – 196.
[20] For definitions of terms related to the description of chiral and achiral crystal structures, see: H. D. Flack, *Helv. Chim. Acta* **2003**, *86*, 905 – 921.
[21] M. R. Sawaya, B. L. Pentelute, S. B. H. Kent, T. O. Yeates, *Acta Crystallogr. Sect. D* **2012**, *68*, 62 – 68.

