Glycoprotein X-ray Structure

DOI: 10.1002/ange.201400679

(Quasi-)Racemic X-ray Structures of Glycosylated and Non-Glycosylated Forms of the Chemokine Ser-CCL1 Prepared by Total Chemical Synthesis**

Ryo Okamoto, Kalyaneswar Mandal, Michael R. Sawaya, Yasuhiro Kajihara, Todd O. Yeates, and Stephen B. H. Kent*

Abstract: *Our goal was to obtain the X-ray crystal structure of the glycosylated chemokine Ser-CCL1. Glycoproteins can be hard to crystallize because of the heterogeneity of the oligosaccharide (glycan) moiety. We used glycosylated Ser-CCL1 that had been prepared by total chemical synthesis as a homogeneous compound containing an N-linked asialo biantennary nonasaccharide glycan moiety of defined covalent structure. Facile crystal formation occurred from a quasi-racemic mixture consisting of glycosylated* L*-protein and non-glycosylated-*D*-protein, while no crystals were obtained from the glycosylated* L*-protein alone. The structure was solved at a resolution of 2.6–2.1 Å. However, the glycan moiety was disordered: only the N-linked GlcNAc sugar was well-defined in the electron density map. A racemic mixture of the protein enantiomers* L*-Ser-CCL1 and* D*-Ser-CCL1 was also crystallized, and the structure of the true racemate was solved at a resolution of 2.7–2.15 Å. Superimposition of the structures of the protein moieties of* L*-Ser-CCL1 and glycosylated-*L*-Ser-CCL1 revealed there was no significant alteration of the protein structure by N-glycosylation.*

[*] Prof. Dr. R. Okamoto,[+] Dr. K. Mandal, Prof. Dr. S. B. H. Kent
Departments of Chemistry: Biochemistry & Molecular Biology
Institute for Biophysical Dynamics, University of Chicago
Chicago, IL 60637 (USA)
E-mail: skent@uchicago.edu

Dr. M. R. Sawaya, Prof. Dr. T. O. Yeates
Department of Chemistry and Biochemistry
University of California
Los Angeles, CA 90095 (USA)

Prof. Dr. Y. Kajihara
Department of Chemistry, Graduate School of Science
Osaka University
Toyonaka, Osaka 5600043 (Japan)

[+] Current address: Department of Chemistry
Graduate School of Science, Osaka University, Toyonaka
Osaka, 560-0043, JAPAN.
E-mail: rokamoto@chem.sci.osaka-u.ac.jp
rokamoto@chem.sci.osaka-u.ac.jp

[**] R.O. gratefully acknowledges receipt of a JSPS Postdoctoral Fellowship for Research Abroad. Use of NE-CAT beamline 24-ID at the Advanced Photon Source is supported by award RR-15301 from the National Center for Research Resources at the National Institutes of Health. Use of the Advanced Photon Source is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201400679.

Glycosylation is one of the most frequent post-translational modifications of proteins in biological systems.[1] It is estimated that over 50% of eukaryotic proteins are glycosylated and three quarters of those glycoproteins are N-glycosylated.[2] This post-translational modification impacts protein properties such as protein folding, immunogenicity, or inflammation.[1] It is well known that the folded, tertiary structure of a protein molecule gives rise to its biological function. Nevertheless, only limited progress has been made in the structural analysis of glycoproteins because of the inherent structural complexity and frequent heterogeneity of the glycan moieties.

X-ray crystallography is a powerful method for the study of protein structures. However, glycoproteins produced by recombinant methods typically have heterogeneous glycan structures and are often hard to crystallize.[3–5] The flexibility of the glycan moiety may also hamper uniform crystal packing. Consequently, the number of reported glycoprotein structures is relatively low. In such a context, we set out to explore a novel methodology for glycoprotein X-ray crystallography. We envisioned preparation of a glycosylated L-protein of defined covalent structure by total chemical synthesis, and facilitated quasi-racemic crystallization of that glycosylated L-protein with the non-glycosylated mirror image D-protein, as a method to obtain diffraction-quality glycoprotein crystals.[6,7]

Chemical protein synthesis is a powerful technique for the preparation of homogeneous post-translationally modified proteins such as glycoproteins, in high purity and with defined chemical structure.[8,9] Furthermore, total chemical synthesis uniquely enables the synthesis of the mirror image D-protein form of a protein molecule.[6,7] We previously demonstrated that a racemic protein mixture composed of equimolar amounts of a naturally occurring L-protein and the mirror image D-protein gave facile crystallization and afforded high-quality protein crystals.[6,10–15] We have extended this approach to the facilitated crystallization of quasi-racemic protein mixtures where an analogue of the L-protein is crystallized with the D-protein form of the parent protein molecule.[6,11] For protein analogues with only minor changes in covalent structure from the parent protein molecule, this quasi-racemic crystallization approach has worked well in our hands.

In the work described here, we set out to explore whether such a quasi-racemate approach could be extended to provide facilitated crystallization of a glycoprotein of natural chirality with the non-glycosylated mirror image D-protein. In order to

test our concept, we carried out the total chemical synthesis of the chemokine Ser-CCL1 in glycosylated and non-glycosylated forms. The chemokine CCL1 was originally defined as a naturally occurring 73 amino acid glycosylated protein having one N-glycosylation site.[16,17] Later, a 74 amino acid form of CCL1, which possesses an additional serine residue at the N-terminal (named Ser-CCL1) was also described (Figure 1).[18] The NMR structure of the non-glycosylated

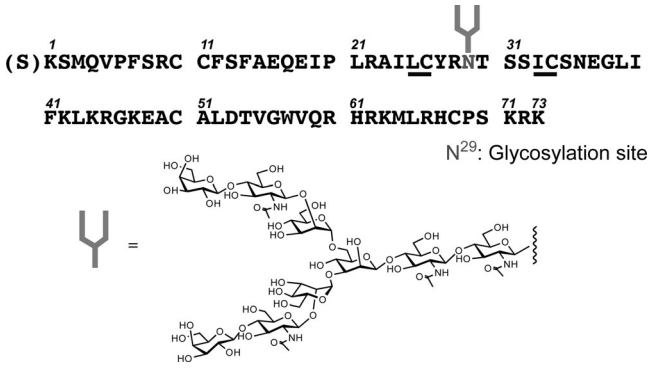

Figure 1. Amino acid sequence of the chemokine Ser-CCL1 and the chemical structure of the N-linked asialo biantennary glycan attached at residue Asn^{29}.

form of Ser-CCL1 has been reported, which showed that this protein fold consists of three antiparallel β-strands and an α-helix. These are the general structural features of chemokines, supported in the case of Ser-CCL1 by an atypical three disulfide bonds; other members of the chemokine family contain just two disulfides.[19] The NMR structure of Ser-CCL1 showed that the N-glycosylation site (Asn^{29}) is at the end of one of the β-strands. This location invoked our curiosity as to whether glycosylation would affect the microstructural environment and perhaps alter the whole structure of the Ser-CCL1 protein molecule.

Here, we describe the analysis of a structurally defined glycosylated form of the chemokine Ser-CCL1. We have previously reported the total chemical synthesis and biological activities of N-glycosylated Ser-CCL1 and non-glycosylated Ser-CCL1 using a modular synthetic strategy.[20] Using the same synthetic strategy, we carried out the total chemical synthesis of the non-glycosylated mirror image protein D-Ser-CCL1, and characterized the structure of the glycoprotein by quasi-racemic crystallization of a mixture comprising glycosylated-L-Ser-CCL1 and D-Ser-CCL1. Diffraction-quality crystals of the glycoprotein quasi-racemate were obtained, while crystals were not obtained from the glycoprotein alone. To the best of our knowledge, this is the first report of the crystal structure of a homogeneous glycoprotein of defined chemical structure prepared by total chemical synthesis.

Total chemical synthesis of the mirror image protein D-Ser-CCL1 was carried out using essentially the same procedure as previously reported (for details, see Supporting Information).[20] Synthetic D-Ser-CCL1 was characterized by liquid chromatography–mass spectrometry (LCMS) and gave an observed mass of 8571.2 ± 0.5 Da in good agreement with the calculated mass of 8571.1 Da (based on average isotope composition).

For crystallization of Ser-CCL1 and glycosylated Ser-CCL1 we prepared three kinds of protein solutions in water: a racemic Ser-CCL1 solution, composed of equimolar amounts of L-Ser-CCL1 and D-Ser-CCL1 (each 5 mg mL^{-1}); glycosylated L-Ser-CCL1 alone (10 mg mL^{-1}); and a quasi-racemic Ser-CCL1 solution, composed of equimolar amounts of glycosylated L-Ser-CCL1 and D-Ser-CCL1 (5.95 mg mL^{-1} and 5 mg mL^{-1}, respectively). Initial screenings were conducted using the 96 conditions of the commercially available Hampton Research Index. Crystals were not obtained from the glycosylated L-Ser-CCL1 alone, even after six months. As we had hoped, we successfully obtained crystals from the quasi-racemic solution of glycosylated L-Ser-CCL1 and D-Ser-CCL1, and also from the racemic Ser-CCL1 solution. Interestingly, the conditions that gave rise to crystals were the same for the quasi-racemate and racemate in the initial screening. The shapes of the quasi-racemate and racemate crystals were different; crystals of the quasi-racemate exhibited only thin needle-like shapes, while the racemate exhibited larger crystals (Supporting Information Figure S7). Single crystals of each type were used to acquire X-ray diffraction data at the Advanced Photon Source (Argonne National Laboratory).

The crystal structure of the true racemate was determined by the molecular replacement method using the program PHASER[21] with the structure of the chemokine CCL14, PDB entry 2Q8T[22] as a search model. The crystal structure of the quasi-racemic mixture was determined by the molecular replacement method using the program PHASER[21] and a search model taken from the crystal structure of the racemic mixture, both L- and D-models. Data collection and refinement statistics are reported in Table S-1 in the Supporting Information. The coordinates of the final models and structure factors have been deposited in the Protein Data Bank with PDB codes 4OIJ (Ser-CCL1 racemate) and 4OIK (Ser-CCL1 quasi-racemate).

The X-ray structure of the Ser-CCL1 true racemate was solved in space group *P*1 and is shown in Figure 2. Structural features consistent with the reported NMR structure[19] of Ser-CCL1 were observed: three antiparallel β-strands and an α-helix, with three disulfide bonds (Figure 2b). It is already known that some chemokines require dimerization to exhibit their function in vivo.[23,24] In the case of non-glycosylated CCL1, it has been shown that this protein does not dimerize under physiological conditions; this was supported by an NMR study at a protein concentration of 2 mM (PDB code 1EL0).[19] In contrast, our crystal structure of non-glycosylated Ser-CCL1 showed a mode of dimerization of the individual enantiomers in the unit cell through an intermolecular antiparallel β-sheet involving the flexible N-terminal region of the polypeptide chain (Figure 2c). This mode of dimerization is in agreement with the typical mode of CC-chemokine dimer formation through the flexible N-terminal region of the polypeptide chain by formation of an intermolecular antiparallel β-sheet.[22] To the best of our knowledge this is the first reported example of a CCL1 dimer structure. In addition, in the unit cell the mirror image L- and D-protein molecules appear to interact through both the N-terminal and C-terminal regions of their polypeptide chains (Figure 2c).

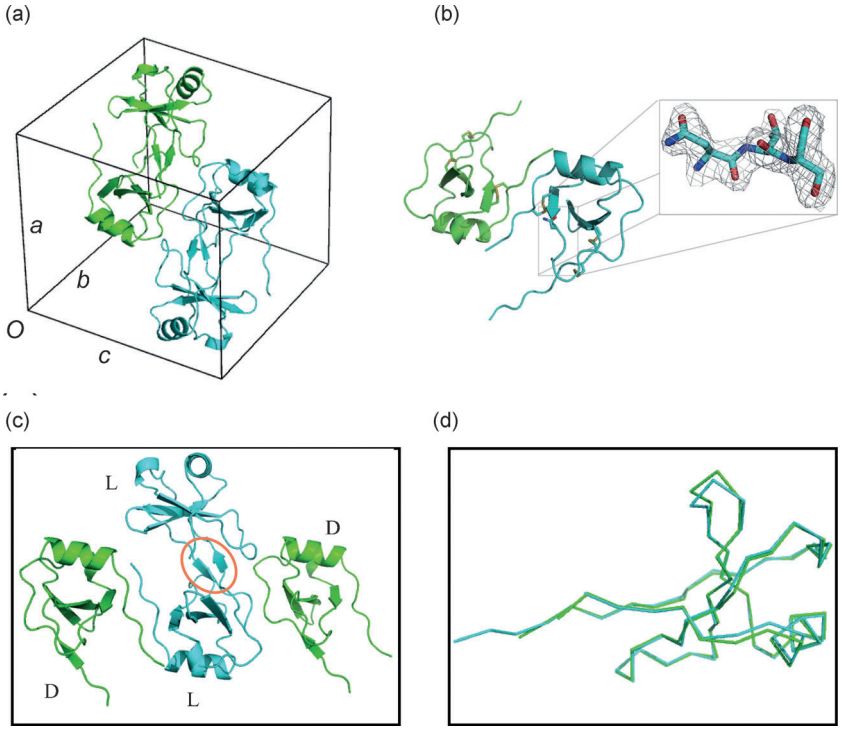

Figure 2. Racemic crystallography of (L-Ser-CCL1 + D-Ser-CCL1). a) Unit cell of the true racemate in space group *P*1 with ribbon representations of the protein molecules. b) Ribbon representation of L-Ser-CCL1 (cyan) and D-Ser-CCL1 (green); Inset: $2F_o-F_c$ electron density map of the (non-glycosylated) Asn^{29}–Ser^{31} consensus sequence for glycosylation shown contoured at a σ level of 1.0. c) D–L interaction interfaces: each protein enantiomer interacts with another molecule of the same chirality forming a dimer through an antiparallel β-sheet in one interface, and interacts with two molecules of opposite chirality involving two other interfaces. The L-molecules are shown in cyan and the D-molecules are in green. The antiparallel β-sheet at the dimer interface is indicated by a red circle. d) Superposition of the X-ray structure of L-protein (cyan) and the inverted D-protein (green) molecules of the protein racemate.

The X-ray structure of the quasi-racemate (glycosylated L-Ser-CCL1 + D-Ser-CCL1) was solved in space group *P*1, and the arrangement of molecules in the unit cell (Figure 3) is quite similar to that observed for the true racemate. We observed all of the structural features typical of the Ser-CCL1 protein molecule: three antiparallel β-strands, an α-helix, and three disulfide bonds.[19] In the quasi-racemate we also observed apparent dimerization of each enantiomer, as was the case for racemic Ser-CCL1 (compare Figure 2a,c, and Figure 3a). Separately for each enantiomer, D-Ser-CCL1 and glycosylated L-Ser-CCL1, the chemokine protein molecules are arranged as dimers formed by interaction of the N-terminal regions of the polypeptide chains. In addition, as for the true racemate, in the unit cell the glycosylated L-Ser-CCL1 and the non-glycosylated D-Ser-CCL1 protein molecules appear to interact using two different interfaces involving both the N-terminal and C-terminal regions of their polypeptide chains. Despite the substantial difference in covalent structures of the D-Ser-CCL1 and glycosylated L-Ser-CCL1 molecules, the enantiomeric protein molecules form a pseudo-centrosymmetrical arrangement within the unit cell.

It has been reported that N-glycosylation could induce a β-turn structure in the polypeptide chain adjacent to the glycosylation site through steric interactions.[3,25] In the quasi-racemic structure of glycosylated Ser-CCL1 we did not observe β-turn-type hydrogen bonding between residues Tyr27 to Thr30, though the GlcNAc residue is visible in the electron density map (Figure 3b, inset). It appeared that this sugar residue is constrained by some steric interactions with neighboring regions of the protein molecule. Glycosylation did not affect the structure of the protein moiety in the glycoprotein molecule as judged by a comparison with the structure of the protein from racemic Ser-CCL1 (Figure 4). The same comparison between glycosylated and

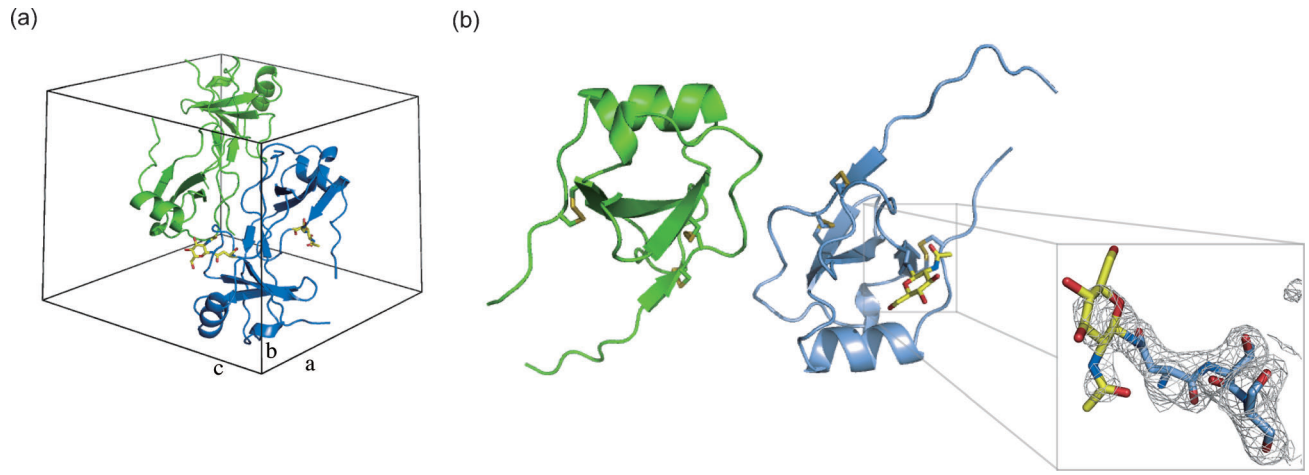

Figure 3. Quasi-racemic crystallography of (glycosylated L-Ser-CCL1 + D-Ser-CCL1). a) Unit cell of the quasi-racemate shown in space group *P*1 with ribbon representation of the protein molecules. b) Ribbon representation of glycosylated L-Ser-CCL1 (light blue) and D-Ser-CCL1 (green); Inset: $2F_o-F_c$ electron density map of the glycosylated Asn^{29}–Ser^{31} consensus sequence is shown contoured at a σ level of 1.0.

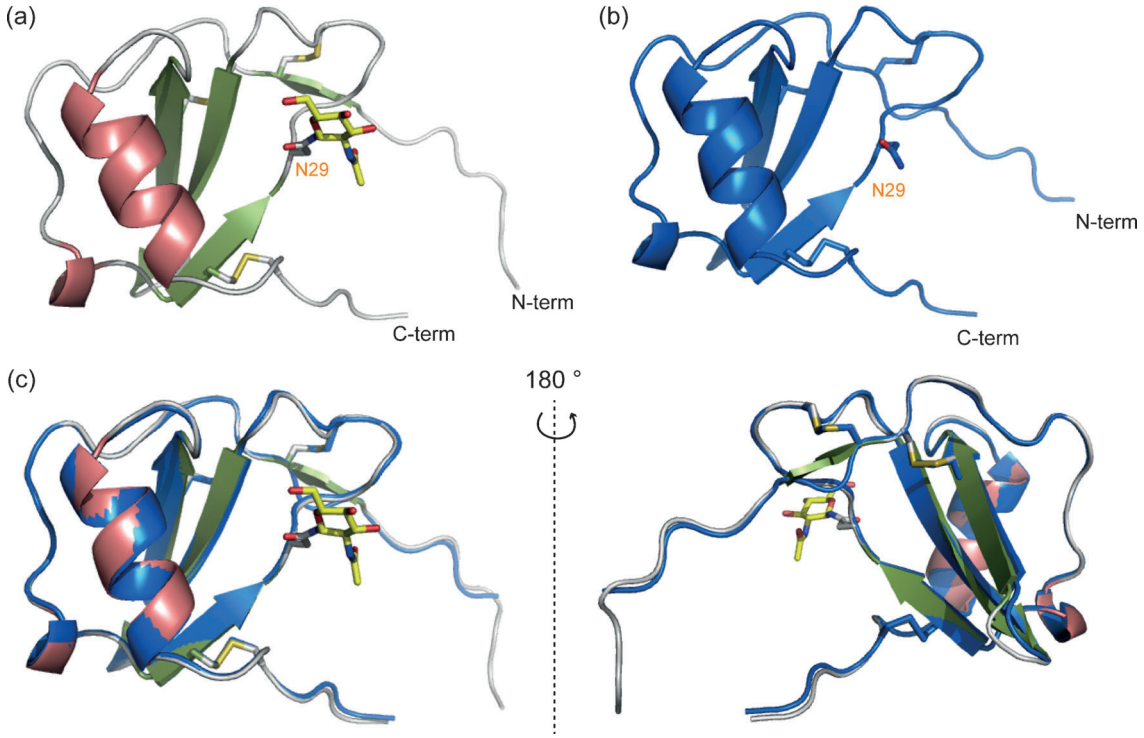

Figure 4. Comparison of the tertiary protein structures of glycosylated and non-glycosylated forms of L-Ser-CCL1. a) X-ray structure of glycosylated L-Ser-CCL1 from the quasi-racemate. b) X-ray structure of non-glycosylated Ser-CCL1 from the true racemate. c) Superimposed structures of glycosylated and non-glycosylated forms of L-Ser-CCL1, shown in two orientations.

non-glycosylated forms of the protein molecule was made using data from the quasi-racemate crystal alone: in this case, the non-glycosylated D-protein Ser-CCL1 structure was computationally inverted and then served as a built-in reference structure for comparison with the protein moiety of the glycoprotein. Both comparisons showed that glycosylation does not perturb the folded structure of the Ser-CCL1 protein molecule.

We had hoped that the facilitated crystallization of glycosylated L-Ser-CCL1 as the quasi-racemate with D-Ser-CCL1 would also result in observation of the complete structure of the complex glycan moiety on the glycoprotein molecule. However, our hopes were not realized: as is typically the case for glycoprotein crystal structures, we could not observe unambiguous electron density for the extended sugar residues on the glycan and could observe only the GlcNAc residue at the reducing end, which is N-linked to Asn^{29} in the Ser-CCL1 protein molecule. Thus, even in the quasi-racemate crystal lattice the glycan moiety is not constrained to one conformation by interactions with the protein molecules, and still exhibited conformational disorder reflective of inherent glycan mobility.

In conclusion, we have determined the X-ray crystal structure of a homogeneous (i.e. a pure, single-molecular species) glycoprotein molecule of defined covalent structure prepared by total chemical synthesis. Diffraction-quality crystals were obtained in a straightforward fashion from a quasi-racemic protein solution, which consisted of equimolar amounts of N-glycosylated L-Ser-CCL1 protein and non-glycosylated D-Ser-CCL1 protein. We also crystallized the true racemate (L-Ser-CCL1 + D-Ser-CCL1). X-ray diffraction data were acquired to 2.15 Å resolution for both crystal forms, and the structures were solved by molecular replacement. Comparison of the crystal structure of glycosylated Ser-CCL1 from the quasi-racemate with non-glycosylated Ser-CCL1 from the true racemate showed that N-glycosylation did not cause significant structural alteration to the tertiary structure of the Ser-CCL1 protein molecule.

While only a limited portion of the glycan moiety could be resolved in this study, we emphasize that the quasi-racemic crystallography approach did make it possible to crystallize the glycosylated protein molecule and to elucidate the structure of the protein component, including the context of its glycosylation. This quasi-racemic crystallization strategy is likely to be useful for other glycoproteins. Furthermore, it might be possible to determine the X-ray structure of the natural glycan moiety in a glycoprotein by quasi-racemic crystallization of the D-glycan-L-protein with an L-neoglycan-D-protein prepared by total chemical synthesis. Such L-neoglycans can be prepared by "click" chemical methods,[26] and their presence in the quasi-racemic crystal lattice may serve to reduce or eliminate the conformational disorder of the glycan moiety observed in the present work, while further facilitating crystallization. We believe that quasi-racemic crystallography of glycoproteins enabled by total chemical synthesis is an important new tool for glycoprotein structural biology.

Received: January 21, 2014
Published online: April 1, 2014

Keywords: chemical protein synthesis · glycoprotein · mirror image proteins · quasi-racemic crystals · X-ray structure

[1] R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683 – 720.
[2] D. Skropeta, *Bioorg. Med. Chem.* **2009**, *17*, 2645 – 2653.
[3] S. E. O'Connor, B. Imperiali, *J. Am. Chem. Soc.* **1997**, *119*, 2295 – 2296.
[4] F. J. López-Jaramillo, F. Pérez-Banderas, F. Hernández-Mateo, F. Santoyo-González, *Acta. Crystallogr. Sect. F* **2005**, *61*, 435 – 438.
[5] C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto, *Angew. Chem.* **2005**, *117*, 7508 – 7539; *Angew. Chem. Int. Ed.* **2005**, *44*, 7342 – 7372.
[6] B. L. Pentelute, Z. P. Gates, V. Tereshko, J. Dashnau, J. M. Vanderkooi, A. A. Kossiakoff, S. B. H. Kent, *J. Am. Chem. Soc.* **2008**, *130*, 9695 – 9701.
[7] T. O. Yeates, S. B. H. Kent, *Annu. Rev. Biophys.* **2012**, *41*, 41 – 61.
[8] S. B. H. Kent, *Chem. Soc. Rev.* **2009**, *38*, 338 – 351.
[9] C. Unverzagt, Y. Kajihara, *Chem. Soc. Rev.* **2013**, *42*, 4408 – 4420.
[10] S. Kent, Y. Sohma, S. Liu, D. Bang, B. Pentelute, K. Mandal, *J. Pept. Sci.* **2012**, *18*, 428 – 436.
[11] K. Mandal, B. L. Pentelute, D. Bang, Z. P. Gates, V. Y. Torbeev, S. B. H. Kent, *Angew. Chem.* **2012**, *124*, 1510 – 1515; *Angew. Chem. Int. Ed.* **2012**, *51*, 1481 – 1486.
[12] K. Mandal, M. Uppalapati, D. Ault-Riché, J. Kenney, J. Lowitz, S. Sidhu, S. B. H. Kent, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14779 – 14784.
[13] M. Avital-Shmilovici, K. Mandal, Z. P. Gates, N. Phillips, M. A. Weiss, S. B. H. Kent, *J. Am. Chem. Soc.* **2013**, *135*, 3173 – 3185.
[14] K. Mandal, B. L. Pentelute, V. Tereshko, A. A. Kossiakoff, S. B. H. Kent, *Protein Sci.* **2009**, *18*, 1146 – 1154.
[15] J. R. Banigan, K. Mandal, M. R. Sawaya, V. Thammavongsa, A. P. Hendrickx, O. Schneewind, T. O. Yeates, S. Kent, *Protein Sci.* **2010**, *19*, 1840 – 1849.
[16] M. D. Miller, S. Hata, R. De Waal Malefyt, M. S. Krangel, *J. Immunol.* **1989**, *143*, 2907 – 2916.
[17] M. D. Miller, M. S. Krangel, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2950 – 2954.
[18] J. M. Fox, P. Najarro, G. L. Smith, S. Struyf, P. Proost, J. E. Pease, *J. Biol. Chem.* **2006**, *281*, 36652 – 36661.
[19] D. W. Keizer, M. P. Crump, T. W. Lee, C. M. Slupsky, I. Clark-Lewis, B. D. Sykes, *Biochemistry* **2000**, *39*, 6053 – 6059.
[20] R. Okamoto, K. Mandal, M. Ling, A. D. Luster, Y. Kajihara, S. B. H. Kent, *Angew. Chem.* **2014**, *126*, DOI: 10.1002/ange.201310574; *Angew. Chem. Int. Ed.* **2014**, *53*, DOI: 10.1002/anie.201310574.
[21] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658 – 674.
[22] K. Y. Blain, W. Kwiatkowski, Q. Zhao, D. La Fleur, C. Naik, T.-W. Chun, T. Tsareva, P. Kanakaraj, M. W. Laird, R. Shah, L. George, I. Sanyal, P. A. Moore, B. Demeler, S. Choe, *Biochemistry* **2007**, *46*, 10008 – 10015.
[23] J. G. Luz, M. Yu, Y. Su, Z. Wu, Z. Zhou, R. Sun, I. A. Wilson, *J. Mol. Biol.* **2005**, *352*, 1019 – 1028.
[24] I. Clark-Lewis, C. Schumacher, M. Baggiolini, B. Moser, *J. Biol. Chem.* **1991**, *266*, 23128 – 23134.
[25] B. Meyer, H. Möller, *Topics in Current Chemistry, Vol. 267* (Ed.: V. Wittmann), Springer, Berlin/Heidelberg, **2007**, pp. 187 – 251.
[26] D. J. Lee, S.-H. Yang, G. M. Williams, M. A. Brimble, *J. Org. Chem.* **2012**, *77*, 7564 – 7571.