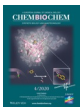
ChemPubSoc Europe

DOI: 10.1002/cbic.201900405

CHEMBIOCHEM Communications



A Tetrahedral Boronic Acid Diester Formed by an Unnatural Amino Acid in the Ligand Pocket of an Engineered Lipocalin

Carina A. Sommer, Andreas Eichinger, and Arne Skerra*[a]

Boronic acids have long been known to form cyclic diesters with *cis*-diol compounds, including many carbohydrates. This phenomenon was previously exploited to create an artificial lectin by incorporating *p*-borono-L-phenylalanine (Bpa) into the ligand pocket of an engineered lipocalin, resulting in a so-called Borocalin. Here we describe the X-ray analysis of its covalent complex with 4-nitrocatechol as a high-affinity model ligand. As expected, the crystal structure reveals the formation of a cyclic diester between the biosynthetic boronate side chain and the two *ortho*-hydroxy substituents of the benzene ring. Interestingly, the boron also has a hydroxide ion associated, despite an only moderately basic pH 8.5 in the crystallization buffer. The complex is stabilized by a polar contact to the side chain of Asn134 within the ligand pocket, thus validating the functional design of the Borocalin as an artificial sugar-binding protein. Our structural analysis demonstrates how a boronate can form a thermodynamically stable diester with a vicinal diol in a tetrahedral configuration in aqueous solution near physiological pH. Moreover, our data provide a basis for the further engineering of the Borocalin with the goal of specific recognition of biologically relevant glycans.

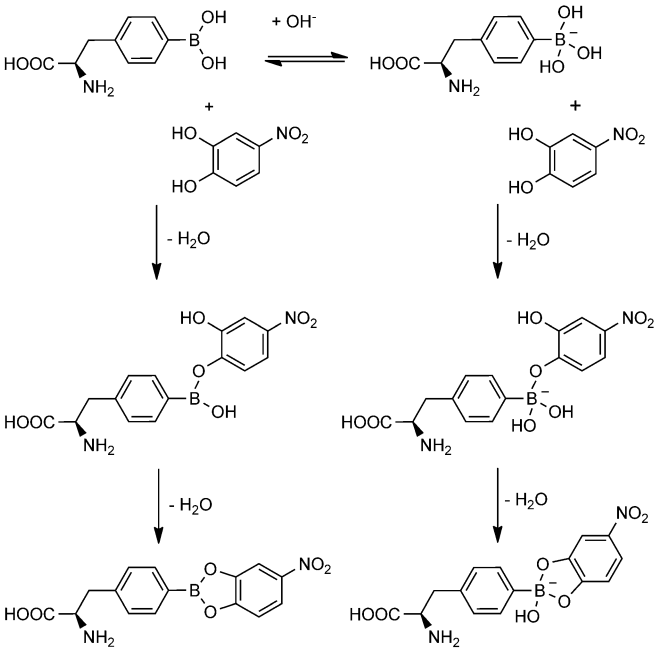
Protein–carbohydrate recognition plays a fundamental role in many biological processes. However, the interactions within individual complexes are relatively weak, with dissociation constants often only in the millimolar range.[1] This generally low affinity of proteins towards sugars can be explained in terms of a camouflage effect of the polyhydroxylated ligands in the aqueous solvent, together with the rotational degrees of freedom of the glycosidic linkages; these lead to multiple conformations of an oligosaccharide[2] and, consequently, to loss of internal entropy upon complex formation. Hence, in order to mediate biological activities at low physiological concentrations, natural lectins use metal ions such as calcium, often together with the avidity effect that arises from multivalent target engagement.

On the other hand, it is well known in chemistry that boronic acids can form covalent diesters with diol groups as they occur, in particular, in carbohydrates.[3] Previously, we functionalized human lipocalin 2 (Lcn2, UniProt ID: P80188) with a boronic acid in its ligand pocket. Therefore, the unnatural amino acid *p*-borono-L-phenylalanine (Bpa) was site-specifically incorporated at position 36 in the recombinant protein through amber stop codon suppression by using *Escherichia coli* cells equipped with an orthogonal BpaRS/tRNA^{UGA} pair.[4] Lipocalins exhibit a robust β-barrel fold with four structurally variable loops at the open end that form a cup-like binding site.[5] In addition to the Bpa residue, three conventional side-chain exchanges were introduced in order to enhance the steric accessibility and chemical reactivity of the boronic acid group [Lcn2(36Bpa-NFW), alias Borocalin], thus providing a basis for the engineering of biomedically relevant sugar specificities by employing Anticalin® technology.[6] For Lcn2(36Bpa-NFW), we have demonstrated binding of pyranose monosaccharides, in particular galactose and mannose, as well as of the aromatic diol 4-nitrocatechol, with dissociation constants in the millimolar and micromolar ranges, respectively.[7]

In its trigonal ground state, boronic acid [$B(OH)_3$, $pK_a \approx 9$], with its unoccupied valence orbital, constitutes a Lewis acid that can readily react with a Lewis base such as OH^- to form a tetrahedral boronate anion, such as in aqueous solution at higher pH.[8] Furthermore, cyclic diesters of boric (or boronic) acids can be formed with organic 1,2 *cis*-diols in a two-step process (Scheme 1), resulting in either the trigonal or the tetrahedral configuration.[3] Apart from the obvious influence of the pH on the acid/base-catalyzed addition/condensation reactions, the tetrahedral cycloadduct is usually preferred for steric reasons because its smaller O–B–O bond angle leads to less ring strain. It is notable that 4-nitrocatechol readily forms much tighter complexes with boronic acid[8] than with pyranose sugars (see above).[7] This could be due to 1) the particular geometry of vicinal OH groups as substituents on a planar unsaturated carbon ring, 2) its aromatic character, and/or 3) the electron-withdrawing and resonance-stabilizing nitro group.[8] The first and second aspect could actually favor the complex in the trigonal boronate configuration, resulting in a fully planar and π-conjugated bicyclic ring system (Scheme 1, left). To investigate whether the biosynthetic boronic acid group within the ligand pocket of the Borocalin would form a mono- or a diester with the hydroxy groups of the 4-nitrocatechol ligand and whether the latter would have a trigonal or tetrahedral configuration (Scheme 1), we elucidated the three-dimensional structure of the complex.

[a] C. A. Sommer, Dr. A. Eichinger, Prof. Dr. A. Skerra
Lehrstuhl für Biologische Chemie, Technische Universität München
Emil-Erlenmeyer-Forum 5, 85354 Freising (Germany)
E-mail: skerra@tum.de

The ORCID identification numbers for the authors of this article can be found under https://doi.org/10.1002/cbic.201900405.



Scheme 1. pH-dependent equilibrium of Bpa, formation of a monoester, and further condensation to an entropically preferred cyclic diester with 4-nitrocatechol in either the trigonal (left) or the tetrahedral (right) configuration.

The crystal structure of the Borocalin with bound 4-nitrocatechol was solved from an X-ray synchrotron data set at 1.98 Å resolution in space group C_2, with two protein molecules in the asymmetric unit, by molecular replacement with the apo-Borocalin[7] as search model. After refinement to a final *R* factor of 20.7% (Table 1), electron density was visible for the two entire polypeptide chains (residues Ser5–Lys188 in chain A, discussed below due to its lower *B* factors) including Bpa at position 36 and, in particular, the covalently bound ligand (Figure 1).

The protein structure of the Borocalin·ligand complex (Figure 1B) appeared virtually unchanged from that of the apo-protein (Figure 1A). Superposition of the structurally conserved set of 58 Cα atoms in the lipocalin fold[5] led to a root mean square deviation (RMSD) as low as 0.225 Å (despite a differing space group). The just slightly higher RMSD of 0.445 Å for all 174 equivalent Cα atoms (Ser5–Gly178, without the Strep-tag II)[10] indicates that, beyond the β-barrel core, the conformations of all loop segments and the α-helix were also conserved, including the set of those four structurally variable loops that shape the ligand pocket at the open end.[5] Only the residues of the C-terminal Strep-tag II peptide (SAWSHPQFEK) were shifted by up to 4.1 Å (at the Cα position), due to an intimate interaction with the neighboring molecule in the crystal lattice in the case of the apo-protein as explained in our previous study.[7]

Table 1. Crystallographic analysis and refinement statistics.[a]

Data collection	
space group	C_2
cell dimensions (*a*, *b*, *c*; α, β, γ) [Å; °]	130.5, 44.2, 92.2; 90.0, 133.5, 90.0
resolution range [Å]	66.89–1.98 (2.08–1.98)
$I/\sigma[I]$	6.8 (2.4)
R_{merge} [%]	7.1 (31.0)
observed/unique reflections	153,049/26,302
completeness	97.5 (98.8)
Refinement statistics	
R_{cryst}/R_{free}	20.7/27.1
$CC_{1/2}$	0.998
protein atoms/water molecules	3056/96
average *B* factor protein/solvent [Å^2]	37.1/38.6

[a] Values in parentheses correspond to the highest-resolution shell.

In the crystallized complex, the artificial Bpa side chain forms a cyclic diester with the 1,2-*cis*-diol groups of 4-nitrocatechol, clearly visible both in the initial, ligand-free electron density and in the finally refined $2F_o-F_c$ map, as well as from a calculated omit density (Figure 1C). Interestingly, the boronate anion unambiguously shows the tetrahedral configuration. Its associated hydroxide ion from the solvent is engaged in a polar contact with the Asn134 side chain (2.7 Å distance, with non-ideal H-bond angle), which appears to stabilize the tetrahedral configuration at the crystallization pH of 8.5 (Figure 1C). It is notable that this side chain is the result of a substitution of the original residue Lys134 in wild-type Lcn2 as part of the design of the Borocalin.[7] The polar interaction causes a slight shift of the Bpa side chain towards Asn134 in relation to the previously described structure of the ligand-free protein.[7] Moreover, the orientation of the 4-nitrocatechol is stabilized by two hydrogen bonds formed between the side chains of Ser68 (2.2 Å distance) and Arg81 (2.9 Å distance) and the nitro group. In particular, Ser68 exhibits a favorable angle geometry between its Cβ and Oγ positions and O1 within the plane of the nitro group (in sp^2 hybridization). Notably, the nitrocatechol ligand also makes contacts to the Strep-tag II at the C terminus of a symmetry-related protein chain that occupies the upper part of the ligand pocket.

The importance of residue Asn134 for the formation of the boronate adduct was experimentally examined by mutation of this position to Ala (36Bpa-AFW). The K_D value of the Borocalin containing the smaller and inert Ala side chain towards the ligand 4-nitrocatechol increased sixfold, to (29.3 ± 0.7) μM, if compared to the Asn134 version [$K_D = (4.8 \pm 0.04)$ μM in a side-by-side measurement; Figure 1D].

Our crystallographic analysis provides general insight into the structure–function relationships of diol recognition by boronate groups. Even though the sugar/diol-binding activity of boric acid and its derivatives has been known for almost a century,[11] structural information on corresponding complexes has remained scarce and was mainly based on conductivity and/or spectroscopic[12] as well as NMR measurements.[13] So far, only a few X-ray structures have been elucidated for cyclic diesters between boronic acids and diol compounds: 1) a small-molecule complex between *m*-aminophenylboronate and catechol [aqua-2-(3-aminophenyl)benzo-1,3,2-dioxaborole][9] and 2) complexes between enzymes and boronic acid inhibitors conjugated to glycerol.[14] In the first case, the boron atom forms part of a five-membered ring annealed to a benzene ring. This generally leads to a lower pK_a; also, the hydroxylated (tetrahedral) boron configuration, with its negative charge, is intramolecu-

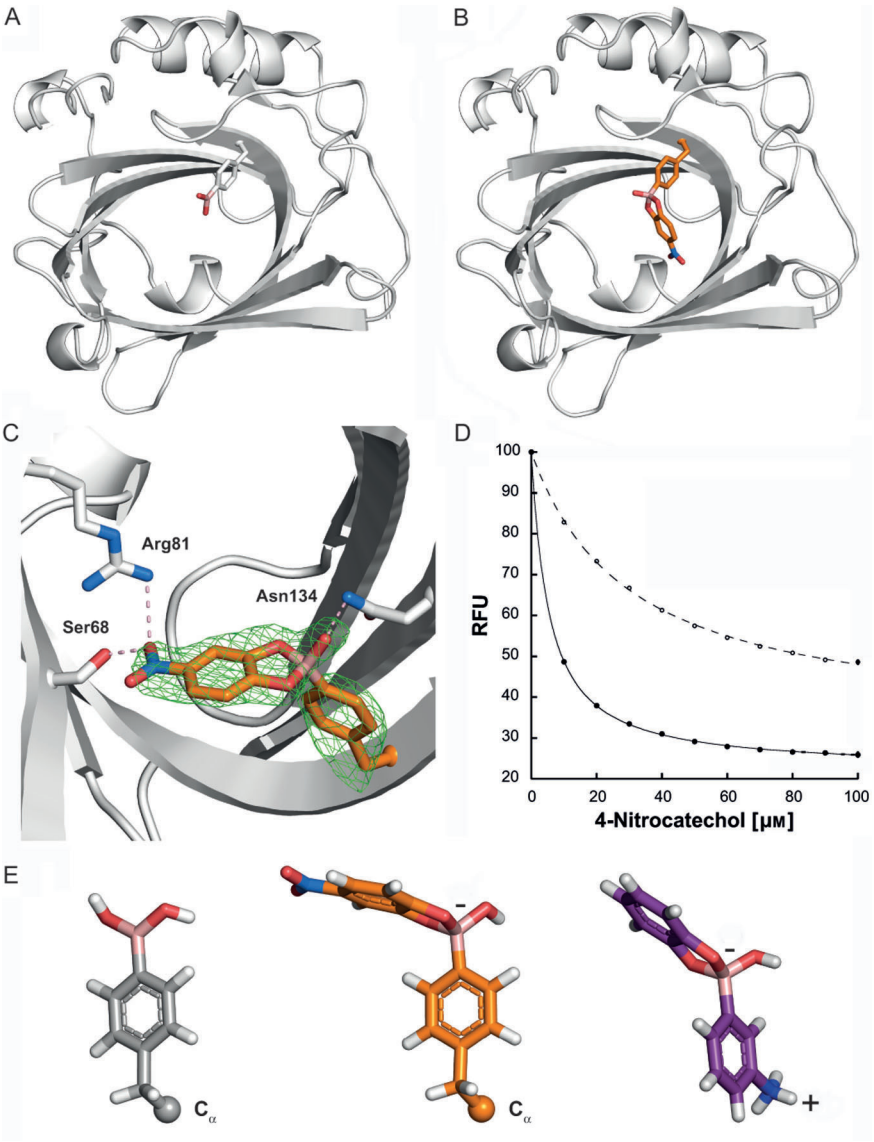


Figure 1. X-ray crystallographic analysis of the Borocalin: A) in the ligand-free state (PDB ID: 5MHH), and B) with covalently bound 4-nitrocatechol (PDB ID: 6QMU). C) Close-up view into the binding pocket (chain A) showing a cyclic diester between the boronic acid and 4-nitrocatechol in the tetrahedral configuration together with an omit electron density map calculated for an Ala side chain (contoured at 2.5 σ). Three H-bonds and/or polar contacts to the ligand are indicated by dashed lines (light pink). D) Fluorescence titration of the Borocalin (36Bpa-NFW: ●, $K_D = 4.8 \pm 0.04\,\mu M$) with 4-nitrocatechol in comparison with the Asn134→Ala mutant (36Bpa-AFW: ○, $K_D = 29.3 \pm 0.7\,\mu M$). E) Structural comparison between the unbound Bpa in the apo-Borocalin (left), the Bpa/4-nitrocatechol diester in the Borocalin ligand pocket (middle), and aqua-2-(3-aminophenyl)benzo-1,3,2-dioxaborole (right).[9] Cα atoms are indicated as spheres; hydrogen positions were modeled for the first two molecules.

larly stabilized by the positive charge of the anilinium group (Figure 1 E, right). In the other examples, glycerol was added at high concentration (typically around 20 %, *v*/*v*) as cryoprotectant during protein crystal handling; hence, glycerol ester formation in the crystal does not indicate a specific high-affinity interaction. In fact, the crystal of the Borocalin·nitrocatechol complex had been treated with glycerol prior to X-ray data collection in the present study, too; however, this did not lead to a detectable ligand exchange.

Thus, the high-resolution crystal structure reported here represents the first example of a specific cyclic diester formed between a solvated boronic acid within the ligand pocket of a protein and a *cis*-diol compound in an aqueous environment (in view of a solvent content of 44.5 % in the protein crystal). The fourth boron ligand position is occupied by a hydroxide anion, which in principle can reversibly dissociate, hence revealing stability of the tetrahedral configuration at pH 8.5 (corresponding to the crystallization buffer for the Borocalin). Our findings should not only support the future development of boronic-acid-based carbohydrate ligands with applications in biological chemistry but also enable further rational reshaping of the ligand pocket of the Borocalin, eventually to obtain a high-affinity binding protein specific for medically relevant cell-surface glycans.

Experimental Section

For X-ray crystallographic analysis, the Borocalin (36Bpa-NFW) was expressed in *E. coli* and purified from the periplasmic extract by using the Strep-tag II essentially as previously described.[7] A protein solution (506 μM) in 2-(*N*-cyclohexylamino)ethanesulfonic acid (CHES, 20 mM)/NaOH (pH 8.5, 330 μL) was mixed at a 1:3 molar ratio with a solution of 4-nitrocatechol in water (10 mM, 50 μL) and crystallized in the presence of ammonium sulfate (1.6 M). Data for a protein crystal, flash-frozen after addition of glycerol (30%, *v*/*v*) to the mother liquor, were collected at BESSY beamline 14.1 at the Helmholtz-Zentrum Berlin, Germany, and processed as described.[7] After molecular replacement (see text), the covalent protein·ligand complex was refined by using B–C (1.610 Å), B–OC (1.520 Å), and B–OH (1.450 Å) bond length parameters from a published small-molecule crystal structure.[9] The simulated annealing omit density map was calculated with PHENIX.[15]

For binding studies, a His_6-tag was employed because in the preceding crystallographic study the C-terminal Strep-tag II had been found to fully occupy the active site of a neighboring Borocalin.[7] To this end, the Borocalin—both Lcn2(36Bpa-NFW) and the variant Lcn2(36Bpa-AFW)—was purified from the *E. coli* periplasmic extract by immobilized metal ion affinity chromatography (IMAC) on a His-Trap HP column (GE Healthcare) with NaH_2PO_4 (40 mM)/NaCl (0.5 M) as running buffer (pH 7.4). The protein was eluted in a linear concentration gradient up to 300 mM imidazole/HCl. Final purification was achieved by cation exchange chromatography on a Resource S column (GE Healthcare) by using 2-morpholin-4-yl-ethanesulfonic acid (MES, 20 mM)/NaOH as buffer (pH 6.0) and elution with a linear concentration gradient up to 0.5 M NaCl. The binding activity for 4-nitrocatechol was determined by fluorescence titration in CHES (50 mM)/NaOH, EDTA (1 mM) pH 8.5 (λ_{ex} = 280 nm, λ_{em} = 345 nm) as previously described.[7]

Acknowledgements

The authors wish to thank Christian Feiler for technical assistance at BESSY beamline 14.1. This work was financially supported by the Helmholtz-Zentrum Berlin, Germany, and the Research Training Group GRK2062 funded by the Deutsche Forschungsgemeinschaft.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: boronic acids · *cis*-diols · cyclic diesters · ligand design · lipocalins · synthetic biology

[1] M. Ambrosi, N. R. Cameron, B. G. Davis, *Org. Biomol. Chem.* **2005**, *3*, 1593 – 1608.
[2] D. B. Walker, G. Joshi, A. P. Davis, *Cell. Mol. Life Sci.* **2009**, *66*, 3177 – 3191.
[3] J. A. Peters, *Coord. Chem. Rev.* **2014**, *268*, 1 – 22.
[4] A. Schiefner, L. Nästle, M. Landgraf, A. J. Reichert, A. Skerra, *Biochemistry* **2018**, *57*, 2597 – 2600.
[5] A. Skerra, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **2000**, *1482*, 337 – 350.
[6] A. Richter, E. Eggenstein, A. Skerra, *FEBS Lett.* **2014**, *588*, 213 – 218.
[7] S. Edwardraja, A. Eichinger, I. Theobald, C. A. Sommer, A. J. Reichert, A. Skerra, *ACS Synth. Biol.* **2017**, *6*, 2241 – 2247.
[8] R. Pizer, L. Babcock, *Inorg. Chem.* **1977**, *16*, 1677 – 1681.
[9] A. Vega, M. Zarate, H. Tlahuext, H. Höpfl, *Z. Kristallogr. New Cryst. Struct.* **2010**, *225*, 676 – 678.
[10] T. G. M. Schmidt, A. Skerra, *Nat. Protoc.* **2007**, *2*, 1528 – 1535.
[11] L. von Vargha, *Ber. Dtsch. Chem. Ges.* **1933**, *66*, 704 – 707.
[12] a) J. Böeseken, *Adv. Carbohyd. Chem.* **1949**, *4*, 189 – 210; b) G. Springsteen, B. Wang, *Tetrahedron* **2002**, *58*, 5291 – 5300.
[13] R. van den Berg, J. A. Peters, H. van Bekkum, *Carbohydr. Res.* **1994**, *253*, 1 – 12.
[14] a) J. Lei, G. Hansen, C. Nitsche, C. D. Klein, L. Zhang, R. Hilgenfeld, *Science* **2016**, *353*, 503 – 505; b) T. I. Lazarova, L. Jin, M. Rynkiewicz, J. C. Gorga, F. Bibbins, H. V. Meyers, R. Babine, J. Strickler, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5022 – 5027; c) C. Nitsche, L. Zhang, L. F. Weigel, J. Schilz, D. Graf, R. Bartenschlager, R. Hilgenfeld, C. D. Klein, *J. Med. Chem.* **2017**, *60*, 511 – 516.
[15] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 213 – 221.

Manuscript received: June 20, 2019
Accepted manuscript online: August 7, 2019
Version of record online: November 7, 2019

