ChemMedChem

Supporting Information

Two Methods, One Goal: Structural Differences between Cocrystallization and Crystal Soaking to Discover Ligand Binding Poses

Barbara Wienen-Schmidt, Matthias Oebbeke, Khang Ngo, Andreas Heine, and Gerhard Klebe*

Table of Contents

1. Crystallographic Tables

Table S1. Crystallographic table for crystal structures obtained by soaking for ligands **1**-**3** and **9**.

Ligand (PDB code)	**9** (6QL0)	**1** (6YNA)	**2** (6YNT)	**3** (6YQK)
Soaked / cocrystallized	**soaked**	**soaked**	**soaked**	**soaked**
Data Collection and Processing				
Wavelength (Å)	0.918409	0.918409	0.918409	0.918409
Beamline	BESSY, 14.1	BESSY, 14.1	BESSY, 14.1	BESSY, 14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M
Space group	$P3_121$ (152)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)
Unit cell parameters:				
a, b, c (Å)	54.7, 54.7, 106.3	58.2, 72.5, 108.9	72.5, 74.5, 80.4	58.2, 73.2, 108.5
α, β, γ (°)	90.0, 90.0, 120.0	90, 90, 90	90, 90, 90	90, 90, 90
Matthews coefficient ($Å^3$/Da)[a]	1.9	2.8	2.64	2.82
Solvent content (%)[a]	36	56.1	53.5	56.3
Diffraction Data[b]				
Resolution range (Å)	47.34-1.63 (1.73-1.63)	45.45-1.47 (1.56-1.47)	43.63-1.52 (1.61-1.52)	45.46-1.67 (1.77-1.67)
Unique reflections	23253 (3510)	79455 (12522)	67316 (10749)	53731 (8549)
Redundancy	7.6 (7.8)	6.7	4.4	5.5
$R(I)_{sym}$ (%)	17.3 (50.4)	3.5 (47.8)	4.4 (44.0)	6.8 (49.0)
Completeness (%)	99.0 (94.3)	99.7 (98.4)	99.5 (99.6)	98.2 (98.1)
<I/σ(I)>	7.2 (2.5)	27.6 (3.8)	17.1 (3.0)	14.3 (2.8)
Refinement				
Resolution range (Å)	47.34-1.63	45.45-1.47	43.63-1.52	43.58-1.67
Reflections (work/free)	22090 / 1163	75482/3973	63950/3366	51044/2687
Final *R* value for all reflections (work/free) (%)	15.8 / 19.5	16.3/18.9	16.9/19.1	16.5/20.2
Protein/peptide residues	223	353/13	346/19	346/13
Calcium atoms	1	0	0	0
Inhibitor atoms	27	20	21	21
Water molecules	165	306	276	240
RMSD from ideality:				
Bond lengths (Å)	0.007	0.005	0.008	0.009
Bond angles (°)	0.9	0.8	1.0	1.0
Ramachandran plot (%)[c]:				
Most favored regions	87.2	93.3	91.3	91.9
Additionally allowed regions	12.8	6.7	8.7	8.1
Generously allowed regions	0.0	0.0	0.0	0.0
Disallowed regions	0.0	0.0	0.0	0.0
Mean *B* factors ($Å^2$)[d]:				
Protein	17.4	27.2	26.4	28.7
Peptide		27.1	24.1	31.4
Ligand	10.9	26.6	30.4	23.0
Water molecules	23.6	34.9	33.1	32.9

[a]Values calculated from Ruppweb[1]
[b]Values in parenthesis refer to the highest resolution shell
[c]Calculated using PROCHECK[2]
[d]Calculated using MOLEMAN excluding hydrogen atoms[3,3,4]

Table S2. Crystallographic table for crystal structures obtained by soaking for ligands **4**-**7**.

Ligand (PDB code)	**4** (6YQJ)	**5** (6YQI)	**6** (6YNB)	**7** (6YNC)
Soaked / cocrystallized	**soaked**	**soaked**	**soaked**	**soaked**
Data Collection and Processing				
Wavelength (Å)	0.918409	0.918409	0.918409	0.918409
Beamline	BESSY, 14.1	BESSY, 14.1	BESSY, 14.1	BESSY, 14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M
Space group	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)
Unit cell parameters:				
a, b, c (Å)	58.2, 72.5, 109.2	58.1, 72.5, 108.7	58.3, 72.0, 109.8	58.3, 73.1, 109.3
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Matthews coefficient ($Å^3$/Da)[a]	2.8	2.8	2.8	2.8
Solvent content (%)[a]	56.2	55.9	56.2	56.1
Diffraction Data[b]				
Resolution range (Å)	45.46-1.58 (1.67-1.58)	45.34-1.42 (1.50-1.42)	43.47-1.72 (1.82-1.72)	45.58-1.40 (1.48-1.40)
Unique reflections	62707 (8401)	86529 (13515)	49381 (7745)	91422 (13988)
Redundancy	5.2	6.7	5.4	6.7
$R(I)_{sym}$ (%)	5.6 (49.5)	3.8 (46.2)	5.1 (49.6)	4.9 (46.7)
Completeness (%)	96.6 (81.1)	98.5 (96.4)	99.3 (97.5)	97.9 (93.7)
<I/σ(I)>	15.67 (2.4)	22.72 (3.1)	20.02 (2.9)	19.01 (3.0)
Refinement				
Resolution range (Å)	45.46-1.58	45.34-1.42	41.81-1.72	45.58-1.40
Reflections (work/free)	59571/3136	82202/4327	46912/2469	86850/4572
Final *R* value for all reflections (work/free) (%)	15.8/20.1	15.4/18.0	15.6/20.1	14.9/17.3
Protein/peptide residues	350/13	347/13	349/13	350/14
Inhibitor atoms	20	20	17	15
Water molecules	342	296	249	372
RMSD from ideality:				
Bond lengths (Å)	0.008	0.007	0.009	0.007
Bond angles (°)	1.0	0.9	1.0	1.0
Ramachandran plot (%)[c]:				
Most favored regions	93.2	93.5	92.2	92.6
Additionally allowed regions	6.8	6.2	7.8	7.1
Generously allowed regions	0.0	0.3	0.0	0.3
Disallowed regions	0.0	0.0	0.0	0.0
Mean *B* factors ($Å^2$)[d]:				
Protein	26.9	27.2	30.6	22.4
Peptide	27.9	30.3	32.1	23.4
Ligand	27.8	27.7	38.4	18.8
Water molecules	34.5	34.8	34.6	31.5

[a]Values calculated from Ruppweb[1]
[b]Values in parenthesis refer to the highest resolution shell
[c]Calculated using PROCHECK[2]
[d]Calculated using MOLEMAN excluding hydrogen atoms[3,3,4]

Table S3. Crystallographic table for crystal structures obtained by co-crystallization for ligands **2** and **8** and by soaking for ligand **8**.

Ligand (PDB code)	**8** (6YNR)	**2** (6Y8C)	**8** (6Y2O)
Soaked / cocrystallized	**soaked**	**co-crystallized**	**co-crystallized**
Data Collection and Processing			
Wavelength (Å)	0.918409	0.970000	0.918409
Beamline	BESSY, 14.1	ESRF, ID29	BESSY, 14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M
Space group	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)
Unit cell parameters:			
a, b, c (Å)	72.4, 74.3, 80.2	58.7, 71.1, 108.0	58.6, 74.9, 106.0
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Matthews coefficient ($Å^3$/Da)[a]	2.6	2.8	2.8
Solvent content (%)[a]	53.3	55.2	56.7
Diffraction Data[b]			
Resolution range (Å)	43.54-1.90 (2.01-1.90)	45.27-1.93 (2.04-1.93)	46.18-2.01 (2.13-2.01)
Unique reflections	34602 (5482)	34803 (5493)	31850 (5005)
Redundancy	4.4	7.4	5.9
$R(I)_{sym}$ (%)	5.6 (49.6)	10.2 (51.8)	5.4 (48.9)
Completeness (%)	99.2 (98.4)	99.8 (99.0)	99.5 (98.4)
<I/σ(I)>	18.8 (3.0)	10.9 (2.8)	19.9 (3.6)
Refinement			
Resolution range (Å)	43.54-1.90	45.27-1.93	46.18-2.01
Reflections (work/free)	32867/1730	34798/1741	30254/1593
Final *R* value for all reflections (work/free) (%)	18.8/22.1	18.5/22.2	22.1/25.4
Protein/peptide residues	341/19	353/18	353/18
Inhibitor atoms	11	21	11
Water molecules	216	162	137
RMSD from ideality:			
Bond lengths (Å)	0.006	0.007	0.007
Bond angles (°)	0.7	0.8	0.8
Ramachandran plot (%)[c]:			
Most favored regions	92.1	92.4	92.4
Additionally allowed regions	7.6	7.6	7.6
Generously allowed regions	0.3	0.0	0.0
Disallowed regions	0.0	0.0	0.0
Mean *B* factors ($Å^2$)[d]:			
Protein	26.6	39.9	37.7
Peptide	24.2	35.7	37.9
Ligand	25.1	38.6	35.3
Water molecules	30.8	41.9	38.5

[a]Values calculated from Ruppweb[1]
[b]Values in parenthesis refer to the highest resolution shell
[c]Calculated using PROCHECK[2]
[d]Calculated using MOLEMAN excluding hydrogen atoms[3,3,4]

All structures not mentioned above can be found in previous publications. This applies for the cocrystallized structures **1**[5],**3**[6], **4**[5] and **5**[5], as well as for **6**[7] and **7**[7]. For the trypsin structure cocrystallized with ligand **9** the crystallographic table can be found in another publication.[8]

2. Comparison of Ligands and Electron Density Maps

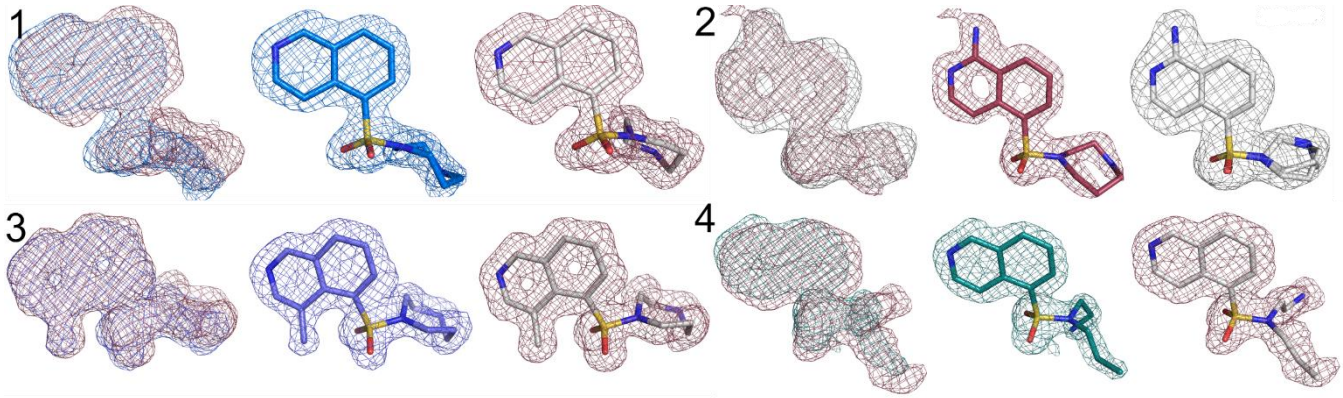

Figure S1. Comparison of the soaked (coloured) and cocrystallized (white) structures 1-4. The $2F_o$-F_c electron densities for the ligands are shown at a RMSD level of 1.0σ except for ligand 2 (0.8σ).

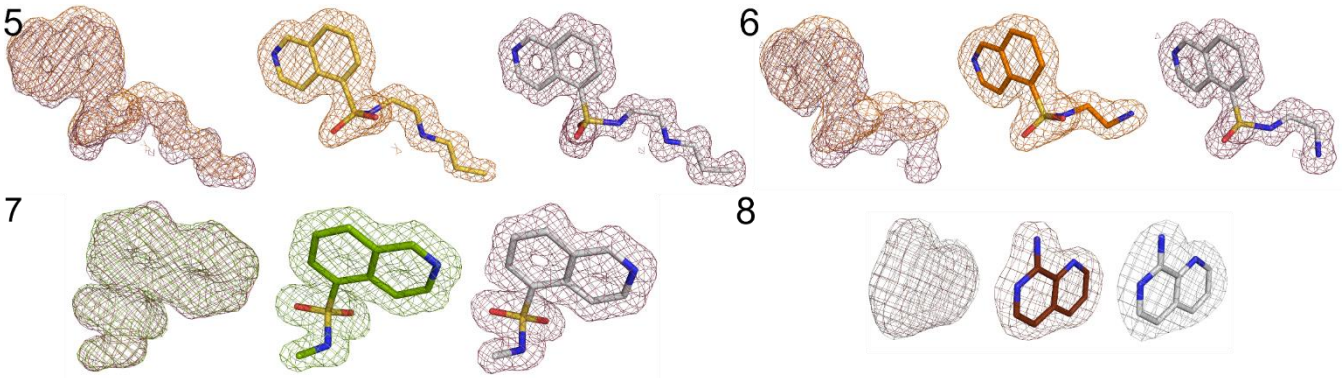

Figure S2. Comparison of the soaked (coloured) and cocrystallized (white) structures 5-8. The $2F_o$-F_c electron densities for the ligands are shown at a RMSD level of 1.0σ.

3. RMSD Calculations

RMSD calculations were performed using ProFitV3.3[9] after fitting the structures with the implemented McLachlan algorithm.[10]

Table S4. RMSD comparison of the cocrystal structures with structures obtained from soaking.

Ligand ID	RMSD C_α [Å][a)]	RMSD C_α HINGE [Å][b)]	RMSD Ligand [Å][c)]	ΔRMSD (Ligand-HINGE) [Å][d)]
1	1.0	0.5	1.0	0.5
2 (1. ligand)	0.8	0.4	0.7	0.3
2 (2. ligand)	0.8	0.4	0.6	0.2
3	0.7	0.3	0.7	0.4
4	1.0	0.4	0.9	0.5
5	0.7	0.4	0.9	0.5
6	0.7	0.3	1.2	0.9
7	0.5	0.2	0.2	0.0
8	0.4	0.2	0.1	0.1

a) RMSD of the fitted C_α atoms based on the fit described in the experimental section
b) RMSD of the C_α atoms 120-123 based on the fit described in the experimental section
c) RMSD of the ligand atoms based on the fit described in the experimental section
d) RMSD difference C_α atoms hinge atoms (b)) and ligand atoms (c))

Table S5. RMSD comparison of all structures using two sets of reference coordinates, one for the cocrystal structures and one for the soaked structures.

	Cocrystallization	Soaking
Ligand ID	**RMSD C_α [Å]** [a)]	**RMSD C_α [Å]** [a)]
1	0.5	0.2
2	0.4	0.3
3	0.5	0.4
4	0.5	0.2
5	0.3	0.2
6	0.3	0.2
7	0.5	0.2
8	0.5	0.3

[a)] The coordinate set was generated using the *MULTREF* command in ProFitV3.3[9] to obtain a mean reference structure to which the overall deviations of the ligand atoms are related. Larger values are computed for the cocrystallized compared to soaked structures.

4. Luzzati Plot

A Luzzati Plot was created and evaluated using the *sfcheck* program from CCP4 to assess the reliability of the crystal structures. The crystal structure with the highest, the lowest and another structure were evaluated (see Table S6). Nevertheless this error does not estimate the final coordinate errors as described by CRUICKSHANK et al..[11]

Table S6. Value of atomic coordinate error by Luzzati Plot listed with the resolution of the crystal structure.

Ligand ID	Resolution [Å]	error by Luzzati Plot
3 (co-crystallized)	1.37	0.2
2 (soaked)	1.52	0.2
8 (co-crystallized)	2.01	0.3

5. Soaking protocol to obtain trypsin crystals of 9

The hanging-drop vapor diffusion method was applied to grow trigonal trypsin crystals at 18°C. For the trypsin-solution a HEPES buffer 15 mM (pH 7.0) with $CaCl_2$ 7.5 mM at a protein concentration of 45 mg mL^{-1} was used. To 1 µL of this solution 1 µL of reservoir buffer was added on a cover slide. The reservoir solution consisted of 0.1 M $(NH_4)_2SO_4$, HEPES 0.1 M (pH 7.5), 25% (w/v) ethylene glycol and 15% (w/v) PEG8000. After seven days of growing well diffracting crystals were obtained, which were also suitable for soaking. Therefore the reservoir buffer was combined with 10% glycerol and 10% ligand solution (1 M in DMSO). At 18°C the crystals were soaked for 5 minutes and flash-frozen in liquid nitrogen.

Data collection, data reduction and structural refinement were performed similar to the protocol described in reference 25 of the main manuscript.

6. References

[1] Bernhard Rupp, "Matthews Probability Calculator", can be found under http://www.ruppweb.org/mattprob/default.html, **2018**.

[2] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J Appl Crystallography* **1993**, *26*, 283.

[3] G. J. Kleywegt, J.-Y. Zou, M. Kjeldgaard, T. A. Jones in *International Tables for Crystallography* (Ed.: H. Fuess), Springer Netherlands, Amsterdam, **2006**, pp. 353–356.

[4] G. J. Kleywegt, "MOLEMAN. Unpublished Program; Uppsala University: Uppsala, Sweden".

[5] B. Wienen-Schmidt, H. R. A. Jonker, T. Wulsdorf, H.-D. Gerber, K. Saxena, D. Kudlinzki, S. Sreeramulu, G. Parigi, C. Luchinat, A. Heine et al., *Journal of medicinal chemistry* **2018**, *61*, 5922.

[6] B. Wienen-Schmidt, D. Schmidt, H.-D. Gerber, A. Heine, H. Gohlke, G. Klebe, *ACS chemical biology* **2019**, *14*, 2585.

[7] B. Wienen-Schmidt, T. Wulsdorf, H. R. A. Jonker, K. Saxena, D. Kudlinzki, V. Linhard, S. Sreeramulu, A. Heine, H. Schwalbe, G. Klebe, *ChemMedChem* **2018**, *13*, 1988.

[8] K. Ngo, C. Collins-Kautz, S. Gerstenecker, B. Wagner, A. Heine, G. Klebe, *Journal of medicinal chemistry* **2020**, *63*, 3274.

[9] C. P. A.C.R. Martin, "ProFitV3.3", can be found under http://www.bioinf.org.uk/software/profit/.

[10] A. D. McLachlan, *Acta Cryst. A* **1982**, *38*, 871.

[11] D.W.J Cruickshank, *Acta Cryst. A* **1996**, 52, C85.