Supporting Information

Hydrogen Bonding Interactions of Covalently-Bonded Fluorine Atoms: From Crystallographic Data to a New Angular Function in the GRID Force Field.

Emanuele Carosati, Simone Sciabola and Gabriele Cruciani*

Laboratory for Chemometrics and Cheminformatics, Department of Chemistry, University of Perugia, Via Elce di Sotto, 10, I-06123, Perugia, Italy

*Corresponding author:
Gabriele Cruciani
Dipartimento di Chimica
Università di Perugia, Via Elce di Sotto, 10, I-06123 Perugia, Italy
Phone: +39 075 5855550
FAX: +39 075 45646
e-mail: **gabri@chemiome.chm.unipg.it**

Table of Contents:

Table S1. Tabular data for the aliphatic fluorine class.

Table S2. Tabular data for the aromatic fluorine class.

Table S3. Tabular data for the geminal fluorine class.

Table S4. Docking application: results.

Table S5. Set of molecules unsuitable for the docking application.

Table S1. Overview of observed X-ray data for the aliphatic fluorine set. Together with the protein and ligand PDB codes, a brief description of the interaction that occurs is reported: the type of interaction (single, double or bifurcated) are specified together with the residues involved in the hydrogen bonding and the measured geometrical properties, which are reported in terms of distances and angles.

Group Code	Protein Code	Chain	Ligand Code	Donor Atom and Residue	Type of Interaction	Distance D···F (Å)	Angle CFH (°)	Angle DHF (°)	Torsional Angle (°)
L1	1exp	/	GLC-G2F (601-602)	ND2 Asn126	Single	1.9	153	141	82
L1	2xyl	/	XYS-X2F (601-602)	ND2 Asn126	Single	2.5	115	132	109
L2	1tsn	/	UFP-C2F (265-266)	OH Tyr94	Single	2.5	142	129	81
L3	4a3h	A	DCB (1)	ND2 Asn138	Double	2.0	129	167	6
				OE2 Glu228	Double	2.2	135	149	0
L4	6a3h	A	FCT (500)	ND2 Asn138	Single	2.0	134	154	6
L4	5a3h	A	FFC (900)	ND2 Asn138	Single	1.9	136	154	2
L4	1qi2	A	FCT (500)	ND2 Asn138	Single	1.9	139	158	6
L4	1h11	A	FCT (500)	ND2 Asn138	Single	2.0	140	155	2
L5	1bvv	A	XYS-DFX (501-502)	NE Arg112	Single	2.2	113	175	20
L5	1c5i	A	XYS-DFX (501-502)	NE Arg112	Single	2.0	117	169	29
L6	2nlr	A	GLC-GLC-G2F (601-602-603)	ND2 Asn155	Double	2.2	115	143	54
				OE1 Glu120	Double	1.7	113	173	50
L7	1b99	A	FUP (160)	NZ Lys16	Single	2.0	142	122	52
L8	1e0v	A	FFC (701)	ND2 Asn127	Single	2.3	117	144	111
L9	1e0x	A	X2F-XYS (701-702)	ND2 Asn127	Single	2.2	122	151	117
		B	X2F-XYS (701-702)	ND2 Asn127	Single	2.3	123	152	116
L10	1e73	M	G2F (999)	ND2 Asn186	Single	2.1	136	144	113
L11	1e70	M	G2F (999)	ND2 Asn186	Single	2.1	130	146	7
L12	1ga8	A	UPF (300)	ND2 Asn153	Single	2.3	119	136	60
L13	1iew	A	G2F (620)	NH1 Arg158	Single	2.1	113	138	69
L14	1qx1	A	FMF (2003)	OD2 Asp92	Single	1.8	121	167	42

Table S2. Overview of observed X-ray data for the aromatic fluorine set. Together with the protein and ligand PDB codes, a brief description of the interaction that occurs is reported: the type of interaction (single, double or bifurcated) are specified together with the residues involved in the hydrogen bonding and the measured geometrical properties, which are specified in terms of distances and angles.

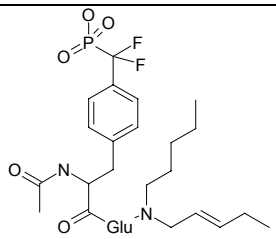
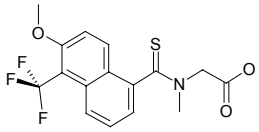
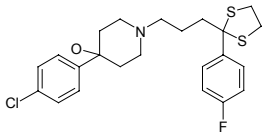
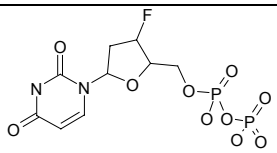
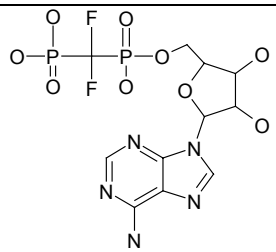
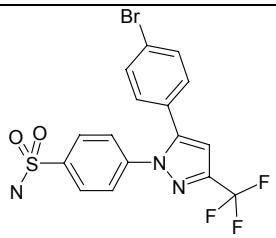
Group Code	protein code	Chain	Ligand Code	Donor Atom and Residue	Type of Interaction	Distance D···F (Å)	Angle CFH (°)	Angle DHF (°)	Torsional Angle (°)
R1	1std	/	BFS (175)	NE2 His110	Single	2.6	150	144	24
R2	1aid	A	THK (201)	OG1 Thr80	Single	2.7	97	142	11
		B	THK (202)	OG1 Thr80	Single	2.3	97	173	24
R3	2aid	A	THK (201)	N Asp30	Single	2.9	130	165	21
R4	1g4j	A	FFB (666)	NE2 Gln92	Single	2.6	117	145	84
R5	1hwi	A	115 (2)	OG Ser661	Single	3.0	91	169	65
		B	115 (1)	OG Ser661	Single	2.8	97	156	70
		C	115 (4)	OG Ser661	Single	3.1	91	164	62
		D	115 (3)	OG Ser661	Single	3.1	90	162	71
R5	1hwj	A	116 (2)	OG Ser661	Single	2.8	99	158	31
		B	116 (1)	OG Ser661	Single	2.9	96	157	20
		C	116 (4)	OG Ser661	Single	3.1	93	167	37
		D	116 (3)	OG Ser661	Single	3.0	95	161	31
R5	1hwk	A	117 (2)	OG Ser661	Single	2.3	107	179	52
		B	117 (1)	OG Ser661	Single	2.3	104	161	37
		C	117 (4)	OG Ser661	Single	2.4	104	179	46
		D	117 (3)	OG Ser661	Single	2.2	108	179	48
R6	1o28	A	UFP (703)	NH1 Arg90	Single	2.9	97	160	8
		B	UFP (708)	NH1 Arg90	Single	2.8	98	160	10
		C	UFP (713)	NH1 Arg90	Single	3.0	93	149	11
		D	UFP (718)	NH1 Arg90	Single	2.8	98	162	7
R6	1o29	A	UFP (803)	NH2 Arg90	Single	2.8	96	161	1
		B	UFP (808)	NH2 Arg90	Single	2.9	94	151	3
		C	UFP (813)	NH2 Arg90	Single	2.8	97	162	0
		D	UFP (818)	NH2 Arg90	Single	2.7	103	168	2

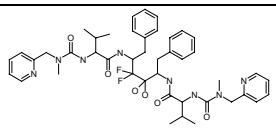
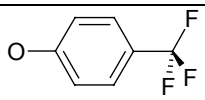
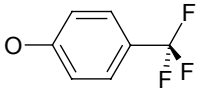
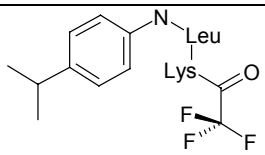
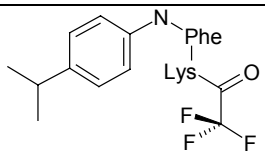
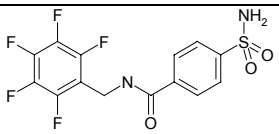
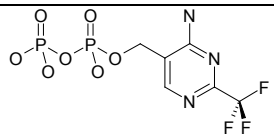
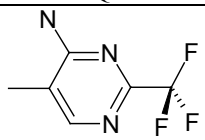
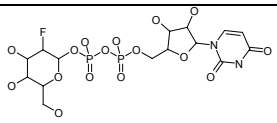
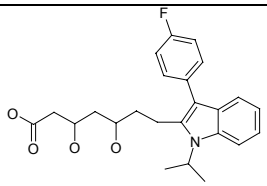
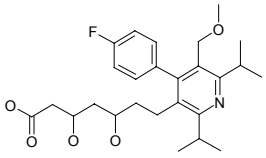
Table S3. Overview of observed X-ray data for the geminal fluorine set. Together with the protein and ligand PDB codes, a brief description of the interaction that occurs is reported: the type of interaction (single, double or bifurcated) are specified together with the residues involved in the hydrogen bonding and the measured geometrical properties, which are reported in terms of distances and angles.

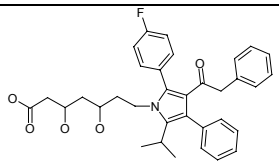
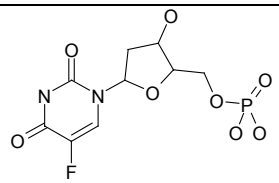
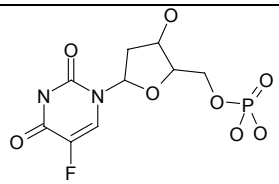
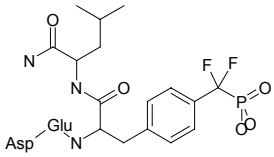
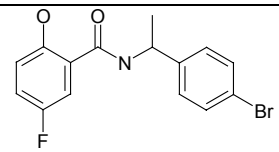
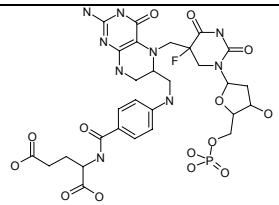
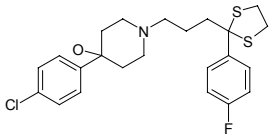
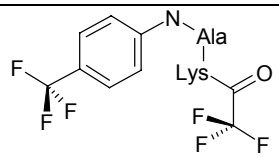
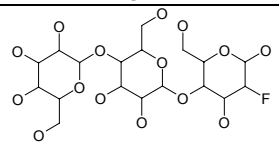
Group Code	protein code	Chain	Ligand Code	Donor Atom and Residue	Type of Interaction	Distance D···F (Å)	Angle CFH (°)	Angle DHF (°)	Torsional Angle (°)
G1	2est	E	TFA-LYS-ALA-ANI(1-4)	OG Ser195	Bifurcated	2.4	71	159	61
				OG Ser195	Bifurcated	2.4	68	147	63
G2	4est	E	ACE-ALA-PRO-VAL-FPA (1-5)	NE2 His57	Single	2.0	101	129	79
G3	6gch	/	APF (246)	NE2 His57	Bifurcated	2.0	138	111	66
				NE2 His57	Bifurcated	3.0	157	62	77
G4	7gch	/	LPF (246)	N Gly193	Single	2.4	98	148	58
G5	1elb	A	TFA-LYS-LEU-ISO (256-259)	N Gly201	Bifurcated	2.5	101	151	70
				N Gly201	Bifurcated	2.8	88	135	39
				OG Ser203	Bifurcated	2.3	90	153	21
				OG Ser203	Bifurcated	2.1	98	147	96
G5	1elc	A	TFA-LYS-PHE-ISO (256-259)	N Gly201	Bifurcated	3.1	83	156	74
				N Gly201	Bifurcated	2.5	110	159	61
				OG Ser203	Bifurcated	2.5	75	137	32
				OG Ser203	Bifurcated	2.0	95	152	76
G6	1dif	A+B[a]	A85 (1)	OD2 Asp25	Bifurcated	2.5	73	148	30
				OD2 Asp25	Bifurcated	1.7	109	151	72
G7	1cx2	A	S58 (701)	NH1 Arg120	Single	2.0	89	173	24
				OH Tyr355	Single	2.5	135	156	119
		B	S58 (701)	NH1 Arg120	Single	2.3	96	169	31
				OH Tyr355	Single	2.6	133	150	13
		C	S58 (701)	NH1 Arg120	Single	1.9	94	175	23
				OH Tyr355	Single	2.3	132	155	5
		D	S58 (701)	NH1 Arg120	Single	2.6	91	170	32
				OH Tyr355	Single	3.1	132	147	18
G8	6cox	A	S58 (701)	NH1 Arg120	Single	2.1	137	161	8
		B	S58 (701)	NH1 Arg120	Single	1.8	137	152	117
G9	1ah3	B	TOL (320)	OG Ser302	Bifurcated	1.9	113	138	77
				OG Ser302	Bifurcated	2.7	76	158	29
G10	6csc	B	COF (700)	N Ala321	Bifurcated	2.4	91	158	43
				N Ala321	Bifurcated	2.5	88	135	72
G11	1a08	A	ACE-FTY-GLU-DIP (100-103)	OG Ser180	Bifurcated	1.8	109	134	26
				OG Ser180	Bifurcated	2.3	86	162	92
		B	ACE-FTY-GLU-DIP (100-103)	OG Ser180	Bifurcated	1.8	104	144	19
				OG Ser180	Bifurcated	2.2	86	149	103
G12	1bwf	O	ATF (601)	OG1 Thr14	Single	2.1	152	129	7
		O	ATF (601)	OG1 Thr267	Bifurcated	2.6	95	149	41
				OG1 Thr267	Bifurcated	2.8	86	161	83
		Y	ATF (601)	OG1 Thr14	Single	2.7	119	160	11
		Y	ATF (601)	OG1 Thr267	Bifurcated	2.0	82	134	90
				OG1 Thr267	Bifurcated	1.8	86	151	30
G13	1dy8	B	CBZ-ILE-LEU-FKI (401-404)	NZ Lys136	Single	2.5	165	139	79
G14	1e0y	B	FCR (601)	OG Ser170	Single	2.4	112	175	118
		B	FCR (601)	OE1 Glu457	Bifurcated	2.2	85	143	82
				OE1 Glu457	Bifurcated	2.0	94	148	23
G15	1e7c	A	HLT (2004)	OG Ser232	Bifurcated	1.9	105	167	119
				OG Ser232	Bifurcated	2.3	85	130	1
G16	1e8g	A	FCR (601)	OG1 Thr457	Single	2.7	116	164	12
		B	FCR (601)	OG1 Thr457	Single	2.9	110	176	17
G17	1g4p	A	FQP (2001)	OG Ser206	Single	1.8	158	180	57
				OH Tyr147	Single	2.7	126	161	13
		B	FQP (2002)	OH Tyr1147	Single	2.7	117	150	22
				OG Ser1206	Single	2.4	163	145	42
G17	1g6c	A	IFP (2001)	OG Ser206	Single	2.4	139	143	42
				OH Tyr147	Single	2.5	126	150	18
		B	IFP (2002)	OH Tyr1147	Single	2.6	124	150	21
				OG Ser1206	Single	2.5	138	141	37
G18	1mmj	N	FR1 (241)	N Gly186	Bifurcated	2.7	156	97	60
				N Gly186	Bifurcated	3.1	161	76	54
G19	1qz0	A	ASP-GLU-FTY-CLE (503-506)	ND2 Gln357	Single	2.1	104	150	45
		B	ASP-GLU-FTY-CLE (503-506)	ND2 Gln357	Single	2.1	103	150	47

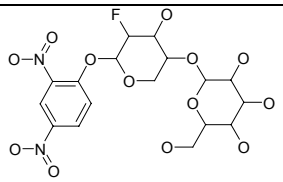
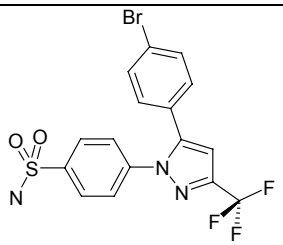
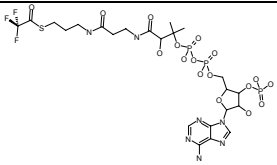
[a] The cavity is composed of residues of both chains A and B, but the ligand is involved with the Asp25 of chain A.

Table S4. For each PDB entry the ligand structure and the values of resolution and R-factor are listed. In addition, the GRID force field without and with the hydrogen bonding capability assigned to the fluorine atoms and described by the new angular function was used in the docking procedure GLUE, to evaluate the effect of fluorine hydrogen bonding capability in protein-ligand recognition phase. Results are reported in terms of rmsd values and ranking of the selected solution referred to the overall number of solutions proposed by the method. The docking procedure was not applied whenever the complex presented covalent bond between ligand and protein.

PDB entry	Res. (Å)	R-factor	Chains[a]	Ligand Structure and Code	No HB			HB		
					rmsd (Å)	Energy (Kcal/mol)	Ranking	rmsd (Å)	Energy (Kcal/mol)	Ranking
1a08	2.20	0.192	A,**B**	 ACE-FTY-GLU-DIP	1.98	-2.0	1/2	0.88	-7.8	1/1
1ah3	2.30	0.207	/	 TOL	0.45	-28.5	1/1	0.39	-29.7	1/1
1aid	2.20	0.174	**A,B**[b]	 THK	0.82	-5.6	10/20	0.80	-5.6	10/20
1b99	2.70	0.207	A	 FUP	1.05	-9.8	1/16	1.03	-9.6	1/17
1bwf	3.00	0.168	**O**,Y	 ATF	0.70	-7.1	1/16	0.70	-9.1	1/20
1cx2	3.00	0.216	A,**B**,C, D	 S58	0.29	-6.3	3/8	0.80	-9.5	1/11

PDB entry	Res. (Å)	R-factor	Chains[a]	Ligand Structure and Code	No HB			HB		
					rmsd (Å)	Energy (Kcal/mol)	Ranking	rmsd (Å)	Energy (Kcal/mol)	Ranking
1dif	1.70	0.198	**A,B**[c]	 A85	0.44	-17.0	2/2	0.44	-17.0	2/2
1e0y	2.75	0.225	B[d]	 FCR	0.79	-8.8	1/18	0.48	-7.8	4/20
1e8g	2.10	0.218	**A**,B	 FCR	0.72	-9.3	1/2	0.82	-9.8	1/2
1elb	2.10	0.150	A	 TFA-LYS-LEU-ISO	1.85	-2.2	20/20	0.56	-7.5	3/20
1elc	1.75	0.150	A	 TFA-LYS-PHE-ISO	/	/	/	1.16	-3.1	12/20
1g4j	1.84	0.183	A	 FFB	/	/	/	1.31	-6.8	2/20
1g4p	2.50	0.219	**A**,B[e]	 FQP	/	/	/	1.12	-7.8	1/20
1g6c	1.40	0.215	**A**,B[f]	 IFP	0.41	-2.1	9/20	0.44	-3.5	3/20
1ga8	2.00	0.189	A	 UPF	0.32	-3.1	1/2	0.36	-3.8	1/2
1hwi	2.30	0.186	**A**,B,C, D	 115	0.35	-11.3	1/1	1.11	-9.0	1/15
1hwj	2.26	0.221	A,**B**,C, D	 116	0.56	-13.6	1/1	0.57	-13.6	1/1

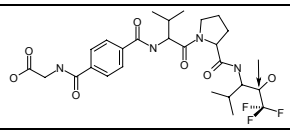
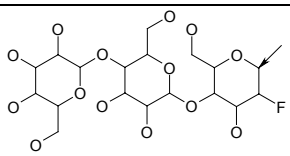
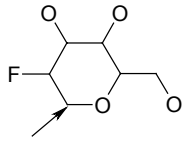
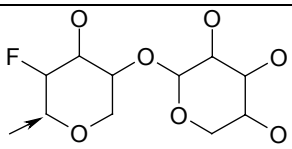
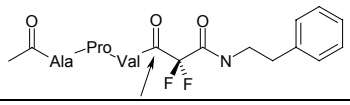
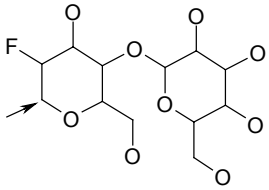
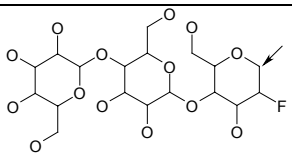
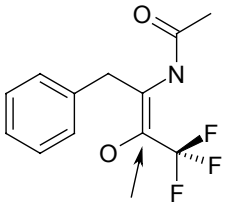
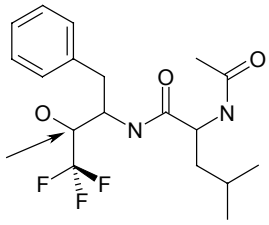
PDB entry	Res. (Å)	R-factor	Chains[a]	Ligand Structure and Code	No HB			HB		
					rmsd (Å)	Energy (Kcal/mol)	Ranking	rmsd (Å)	Energy (Kcal/mol)	Ranking
1hwk	2.22	0.212	**A**,B,C, D	 117	0.39	-14.2	1/1	0.37	-12.6	1/1
1o28	2.10	0.193	**A**,B,C, D[g]	 UFP	0.24	-9.1	1/1	0.40	-7.8	1/6
1o29	2.00	0.194	A,B,**C**, D[h]	 UFP	0.63	-12.5	1/1	0.30	-10.4	1/1
1qz0	1.50	0.188	A,**B**[i]	 ASP-GLU-FTY-CLE	/	/	/	0.26	-11.8	1/18
1std	2.90	0.179	/	 BFS	0.86	-20.1	3/4	0.88	-20.9	3/4
1tsn	2.20	0.178	/	 UFP-C2F	0.91	-5.4	3/3	0.97	-11.6	1/3
2aid	1.90	0.179	**A,B**[c]	 THK	0.96	-4.2	8/20	1.38	-7.2	1/20
2est	2.50	0.210	E	 TFA-LYS-ALA-ANI	0.54	-3.0	5/15	1.15	-6.6	3/20
2nlr	1.20	0.111	A	 GLC-GLC-G2F	0.53	-8.7	7/12	0.27	-10.8	4/15

PDB entry	Res. (Å)	R-factor	Chains[a]	Ligand Structure and Code	No HB			HB		
					rmsd (Å)	Energy (Kcal/mol)	Ranking	rmsd (Å)	Energy (Kcal/mol)	Ranking
4a3h	1.65	n.a.	A	 DCB	0.92	-2.6	12/20	0.38	-2.5	11/20
6cox	3.00	0.220	**A**,B	 S58	0.73	-7.6	1/2	0.91	-7.7	1/2
6csc	2.25	0.158	B[j]	 COF	0.47	-5.9	1/2	0.45	-4.0	1/2

[a] The chain used for the docking is reported in bold. [b] Same binding site for chains A and B. The fourth docking solution of the ligand THK 2001 was posed over the ligand THK 2002. The water molecule 1602 was added within the binding site. [c] Same binding site for chains A and B. [d] No ligand is present in the chain A. [e] Mg 2003 from the chain A was included. [f] POP 2003, TZP 2005 and Mg 2007 from the chain A were included. [g] Ligands EPE and UFP were included. [h] Ligands UFP and the co-factor FAD were included. [i] The ligand FTY was included, and the chains D and F were not considered because of their different binding sites. [j] In the chain A the hydrogen bond was not revealed.

Table S5. Molecules not docked. The reason responsible for the choice of not docking the ligand is reported in the Note column and the line of the PDB file where this reason is proven is reported in brackets.

PDB entry	Ligand Code	Res. (Å)	R-factor	Ligand Structure	Note
1bvv	XYS-DFX	1.80	0.189		Exists covalent bond between C1 of DFX and OE2 of Glu78 (line 278)
1c5i	XYS-DFX	1.80	0.193		Exists covalent bond between C1 of DFX and OE2 of Glu78 (line 322)
1dy8	CBZ-ILE-LEU-FKI	2.40	0.216		Chain A: Not formed hydrogen bond. Exists covalent bond between C of FKI and OG of Ser139 (line 507)
1e0v	FFC	1.70	0.158		Exists covalent bond between C1A of FFC and OE1 of Glu236 (line 359)
1e0x	X2F-XYS	1.65	0.126		Exists covalent bond between C1 of X2F and OE1 of Glu236 (line 500, 502)
1e70	G2F	1.65	0.169		Glu409 is twice defined for atoms CA, CB, CG, CD, whereas OE1 and OE2 are defined only once, OE1 close to Chain B, OE2 close to chain A (line 3849)
1e73	G2F	1.50	0.134		Exists covalent bond between C1 of G2F and OE1 of Glu409 (line 578)
1e7c	HLT	2.40	0.232		7 different binding sites for 7 molecules of the ligand HLT. No specificity.
1exp	GLC-G2F	1.80	0.208		Exists covalent bond between C1 of G2F and OE1 of Glu233 (line 342)
1h11	FCT	1.08	0.108		Exists covalent bond between C1A of FCT and OE2 of Glu228 (line 384)
1iew	G2F	2.55	0.189		Exists covalent bond between C1 of G2F and OD1 of Asp285 (line 470)

PDB entry	Ligand Code	Res. (Å)	R-factor	Ligand Structure	Note
1mmj	FR1	2.20	0.215		Exists covalent bond between C31 of FR1 and OG of Ser188 (line 385)
1qi2	FCT	1.75	0.170		Exists covalent bond between C1A of FCT and OE2 of Glu228 (line 379)
1qx1	FMF	1.30	0.170		Exists covalent bond between C1 of FMF and OD1 of Asp204 (line 602)
2xyl	XYS-X2F	1.90	0.212		Exists covalent bond between C1 of X2F and OE1 of Glu233 (line 314)
4est	ACE-ALA-PRO-VAL-FPA	1.78	0.160		Exists covalent bond between C of Val4 and OG of Ser195 (lines 78-79)
5a3h	FFC	1.82	0.144		Exists covalent bond between C1A of FFC and OE2 of Glu228 (line 316)
6a3h	FCT	1.68	0.150		Exists covalent bond between C1A of FCT and OE2 of Glu228 (line 347)
6gch	APF	2.10	0.180		Exists covalent bond between C2 of APF and OG of Ser195 (lines 3-4)
7gch	LPF	1.80	0.190		Exists covalent bond between C2 of APF and OG of Ser195 (lines 3-4)