Structure *19*

Supplemental Information

Status of GPCR Modeling and Docking as Reflected

by Community-wide GPCR Dock 2010 Assessment

Irina Kufareva, Manuel Rueda, Vsevolod Katritch, GPCR Dock 2010 participants, Raymond C. Stevens, and Ruben Abagyan

Table S1. Protein prediction accuracy in GPCR Dock 2010, related to Figures 1B, 2C, and Table 1.

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
D3/eticlopride									
WUStL	3	1.35(205)	72%(0.76)	2.69(15)	n/a	1.11	ClustalW [1], Modeller v9.2 [2], NAMD 2.6 [3], Rosetta [4]	Alignment refinement	
WUStL	2	1.37(205)	70%(0.75)	2.72(15)	n/a	1.07	ClustalW [1], Modeller v9.2 [2], NAMD 2.6 [3], Rosetta [4]	Alignment refinement	
Monash-Hall	4	1.39(205)	61%(0.53)	2.72(15)	n/a	1.04	ICM [5]	Alignment refinement	
PompeuFabra	3	1.38(205)	63%(0.52)	2.87(15)	n/a	1	Modeller v9.8 [2], MOE [6], PROCHECK [7]	Alignment refinement, Model selection	
PompeuFabra	1	1.38(205)	65%(0.56)	3.(15)	n/a	0.97	Modeller v9.8 [2], MOE [6], PROCHECK [7]	Alignment refinement, Model selection	
PompeuFabra	2	1.38(205)	65%(0.56)	3.01(15)	n/a	0.96	Modeller v9.8 [2], MOE [6], PROCHECK [7]	Alignment refinement, Model selection	
Monash-Hall	1	1.39(205)	61%(0.52)	3.07(15)	n/a	0.93	ICM [5]	Alignment refinement	
Monash-Sexton-2	2	1.37(205)	63%(0.57)	3.31(15)	n/a	0.88	Modeller [2], OPUS-PSP [8], CONCOORD [9], GROMACS [10], RAPPER [11]		
Monash-Hall	3	1.38(205)	63%(0.55)	3.33(15)	n/a	0.87	ICM [5]	Alignment refinement	
WUStL	1	1.44(205)	70%(0.78)	3.11(15)	n/a	0.86	ClustalW [1], Modeller v9.2 [2], NAMD 2.6 [3], Rosetta [4]	Alignment refinement	
KIAS	3	1.48(205)	63%(0.63)	3.(15)	n/a	0.85	LOMETS [12], SPICKER [13], I-TASSER [14], FG-MD [15]		
KIAS	2	1.49(205)	63%(0.71)	2.97(15)	n/a	0.84	LOMETS [12], SPICKER [13], I-TASSER [14], FG-MD [15]		

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
Monash-Hall	5	1.37(205)	65%(0.57)	3.44(15)	n/a	0.84	ICM [5]	Alignment refinement	
UNM	4	1.39(205)	66%(0.59)	3.46(15)	n/a	0.8	3D-Jury [16], Modeller v9.8 [2]	Alignment refinement	
UNM	1,2	1.39(205)	66%(0.58)	3.47(15)	n/a	0.8	3D-Jury [16], Modeller v9.8 [2]	Alignment refinement	
Monash-Hall	2	1.35(205)	61%(0.53)	3.64(15)	n/a	0.79	ICM [5]	Alignment refinement	
UNM	3	1.45(205)	66%(0.57)	3.34(15)	n/a	0.77	3D-Jury [16], Modeller v9.8 [2]	Alignment refinement	
UNM	5	1.46(205)	65%(0.57)	3.34(15)	n/a	0.76	3D-Jury [16], Modeller v9.8 [2]	Alignment refinement	
KIAS	1	1.51(205)	63%(0.71)	3.14(15)	n/a	0.75	LOMETS [12], SPICKER [13], I-TASSER [14], FG-MD [15]		
KIAS	5	1.41(205)	64%(0.58)	3.66(15)	n/a	0.72	LOMETS [12], SPICKER [13], I-TASSER [14], FG-MD [15]		
UMich-Zhang	5	1.46(205)	66%(0.65)	3.47(15)	n/a	0.72	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
UCSF-Shoichet-1	4	1.45(205)	65%(0.6)	3.53(15)	n/a	0.71	PROMALS-3D [17], Modeller v9.0 [2], 3K-ENM [18], DOPE [19]	Alignment refinement	
UMich-Zhang	1	1.5(205)	65%(0.7)	3.35(15)	n/a	0.71	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
UMich-Lomize	1	1.45(205)	63%(0.58)	3.63(15)	n/a	0.68	[20, 21], DIANA [22]	Alignment refinement	
UMich-Pogozheva	1	1.45(205)	63%(0.58)	3.63(15)	n/a	0.68	[20, 21], DIANA [22]	Alignment refinement	
UMich-Zhang	4	1.44(205)	66%(0.65)	3.74(15)	n/a	0.65	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
RHUL	1,2,3,4,5	1.7(205)	65%(0.59)	2.75(15)	n/a	0.64	Modeller [2]		
UNSW	3,4	1.43(205)	62%(0.57)	3.93(15)	n/a	0.6	ClustalW [1], Accelrys Discovery Studio [23]	Alignment refinement	
PharmaDesign	1	1.51(205)	64%(0.56)	3.72(15)	n/a	0.58	ClustalW [1], MOE 2007 [6]	Alignment refinement	
Helsinki-Xhaard	1	1.55(205)	67%(0.67)	3.74(15)	n/a	0.51	Modeller v9.2 [2], SYBYL-X [24]	Side-chain rotamer adjustment	
UMich-Zhang	2	1.5(205)	65%(0.7)	4.09(15)	n/a	0.47	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
UNSW	5	1.63(205)	65%(0.6)	3.63(15)	n/a	0.45	ClustalW [1], SwissModel [25]	Alignment refinement	
KIAS	4	1.56(205)	63%(0.61)	3.96(15)	n/a	0.43	LOMETS [12], SPICKER [13], I-TASSER [14], FG-MD [15]		

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
Baylor-Barth	3	1.71(205)	67%(0.67)	3.42(15)	n/a	0.41	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		
Stockholm	5	1.58(205)	66%(0.58)	3.92(15)	n/a	0.41	Pcons.net [27], Modeller v9.8 [2]		
COH-Vaidehi	3	1.55(205)	71%(0.81)	4.06(15)	n/a	0.4	Modeller [2], LiTiCon [28]	ECL2 and ECL3 conformation adjustment	
COH-Vaidehi	2	1.57(205)	70%(0.8)	4.01(15)	n/a	0.4	Modeller [2], LiTiCon [28]	ECL2 and ECL3 conformation adjustment	
COH-Vaidehi	1	1.58(205)	69%(0.79)	4.11(15)	n/a	0.36	Modeller [2], LiTiCon [28]	ECL2 and ECL3 conformation adjustment	
GaTech	1,2,3,4,5	1.88(205)	69%(0.89)	3.06(15)	n/a	0.31	TASSERmfret [29]		
UMich-Zhang	3	1.26(205)	74%(0.81)	5.54(15)	n/a	0.31	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
Baylor-Barth	1	1.75(205)	64%(0.75)	3.77(15)	n/a	0.24	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		
Monash-Yuriev	5	1.48(205)	63%(0.56)	4.9(15)	n/a	0.23	Prime 2.2 [30]	Alignment refinement	
Monash-Sexton-2	1	1.63(205)	64%(0.69)	4.35(15)	n/a	0.21	Modeller [2], OPUS-PSP [8], CONCOORD [9], GROMACS [10], RAPPER [11]		
Monash-Yuriev	4	1.48(205)	62%(0.55)	4.96(15)	n/a	0.21	Prime 2.2 [30]	Alignment refinement	
Monash-Yuriev	1	1.49(205)	63%(0.56)	4.98(15)	n/a	0.2	Prime 2.2 [30]	Alignment refinement	
Monash-Yuriev	2	1.55(205)	64%(0.64)	4.78(15)	n/a	0.18	Prime 2.2 [30]	Alignment refinement	
UNSW	1,2	1.55(205)	65%(0.57)	4.8(15)	n/a	0.17	ClustalW [1], Accelrys Discovery Studio [23]	Alignment refinement	
Baylor-Barth	4	1.74(205)	64%(0.72)	4.14(15)	n/a	0.14	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		
Monash-Yuriev	3	1.53(205)	62%(0.54)	5.01(15)	n/a	0.13	Prime 2.2 [30]	Alignment refinement	
Stockholm	3	1.56(205)	67%(0.66)	4.89(15)	n/a	0.13	Pcons.net [27], Modeller v9.8 [2]		
Monash-Sexton-2	3	1.61(205)	63%(0.62)	4.81(15)	n/a	0.09	Modeller [2], OPUS-PSP [8], CONCOORD [9], GROMACS [10], RAPPER [11]		
QUB	4	1.71(205)	64%(0.62)	4.45(15)	n/a	0.08	ClustalW [1], MOE [6]	Alignment refinement	
QUB	1	1.72(205)	64%(0.63)	4.46(15)	n/a	0.07	ClustalW [1], MOE [6]	Alignment refinement	

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
QUB	2	1.73(205)	63%(0.62)	4.49(15)	n/a	0.04	ClustalW [1], MOE [6]	Alignment refinement	
QUB	3	1.71(205)	65%(0.64)	4.56(15)	n/a	0.04	ClustalW [1], MOE [6]	Alignment refinement	
Monash-Sexton-2	4	1.56(205)	60%(0.55)	5.18(15)	n/a	0.04	Modeller [2], OPUS-PSP [8], CONCOORD [9], GROMACS [10], RAPPER [11]		
QUB	5	1.73(205)	64%(0.62)	4.54(15)	n/a	0.03	ClustalW [1], MOE [6]	Alignment refinement	
UWash	3	2.05(205)	65%(0.79)	3.28(15)	n/a	0.02	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
UCSF-Shoichet-1	2	1.52(205)	65%(0.63)	5.57(15)	n/a	-0.04	PROMALS-3D [17], Modeller v9.0 [2], 3K-ENM [18], DOPE [19]	Alignment refinement	
Baylor-Barth	2	1.65(205)	67%(0.55)	5.05(15)	n/a	-0.04	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		
Caltech	2	1.77(205)	77%(1.12)	4.63(15)	n/a	-0.05	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
CDD-CMBI	5	1.94(205)	67%(0.67)	4.(15)	n/a	-0.07	Yasara[35]	Model refinement	
Caltech	1	1.97(205)	68%(1.04)	3.99(15)	n/a	-0.1	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
UWash	5	2.09(205)	66%(0.88)	3.57(15)	n/a	-0.12	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
UWash	1	2.03(205)	66%(0.88)	3.81(15)	n/a	-0.12	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
CDD-CMBI	4	1.94(205)	67%(0.66)	4.17(15)	n/a	-0.12	Yasara[35]	Model refinement	
UWash	2	2.17(205)	66%(0.88)	3.27(15)	n/a	-0.13	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
CDD-CMBI	3	1.95(205)	66%(0.65)	4.35(15)	n/a	-0.19	Yasara[35]	Model refinement	
CDD-CMBI	1	1.95(205)	66%(0.66)	4.38(15)	n/a	-0.2	Yasara[35]	Model refinement	
CDD-CMBI	2	1.95(205)	67%(0.66)	4.38(15)	n/a	-0.2	Yasara[35]	Model refinement	
Stockholm	1,2	1.7(205)	66%(0.66)	5.55(15)	n/a	-0.26	Pcons.net [27], Modeller v9.8 [2]		
Monash-Sexton-2	5	1.61(205)	61%(0.56)	6.4(15)	n/a	-0.41	Modeller [2], OPUS-PSP [8], CONCOORD [9], GROMACS [10], RAPPER [11]		
Warsaw	1,2,3,4,5	2.13(205)	62%(0.64)	4.71(15)	n/a	-0.53	MUSCLE[36], ClustalW [1], Modeller [2],	Alignment refinement	

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
Soochow	1	2.33(205)	62%(0.66)	4.14(15)	n/a	-0.61	ClustalW [1], Accelrys Discovery Studio [23], CCBTX [37]		
Stockholm	4	1.56(205)	67%(0.66)	7.39(15)	n/a	-0.67	Pcons.net [27], Modeller v9.8 [2]		
Caltech	5	2.04(205)	66%(0.94)	5.79(15)	n/a	-0.77	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
UNC	1,4,5	2.56(205)	63%(0.65)	3.8(15)	n/a	-0.79	ClustalW [1], Medusa [38], Chiron [39]		Ab-initio sampling of loops
Caltech	4	2.22(205)	64%(1.)	5.24(15)	n/a	-0.82	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
Caltech	3	2.39(205)	66%(1.22)	4.66(15)	n/a	-0.85	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
Evotec	1,2,3,4,5	2.38(205)	74%(1.17)	4.96(15)	n/a	-0.93	MOE [6]		
Baylor-Barth	6	1.81(205)	63%(0.7)	7.42(15)	n/a	-1	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		No ligand
COH-Vaidehi	4	2.68(205)	41%(1.35)	4.37(15)	n/a	-1.13	MembStruk [40], Modeller [2]	ECL2 and ECL3 conformation adjustment	Ab-initio modeling of TM bundle
COH-Vaidehi	5	2.7(205)	41%(1.33)	4.37(15)	n/a	-1.15	MembStruk [40], Modeller [2]	ECL2 and ECL3 conformation adjustment	Ab-initio modeling of TM bundle
UWash	4	1.96(205)	67%(0.83)	7.46(15)	n/a	-1.2	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
UCSF-Shoichet-1	5	1.9(205)	68%(1.05)	7.94(15)	n/a	-1.28	PROMALS-3D [17], Modeller v9.0 [2], 3K-ENM [18], DOPE [19]	Alignment refinement	
Baylor-Barth	5	2.03(205)	66%(0.89)	7.54(15)	n/a	-1.32	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		No ligand
UCSF-Shoichet-1	3	1.98(205)	66%(0.85)	8.16(15)	n/a	-1.46	PROMALS-3D [17], Modeller v9.0 [2], 3K-ENM [18], DOPE [19]	Alignment refinement	
UNC	2,3	2.55(205)	62%(0.61)	7.3(15)	n/a	-1.9	ClustalW [1], Medusa [38], Chiron [39]		Ab-initio sampling of loops
UCSF-Shoichet-1	1	2.25(205)	65%(0.88)	8.88(15)	n/a	-2.03	PROMALS-3D [17], Modeller v9.0 [2], 3K-ENM [18], DOPE [19]	Alignment refinement	
Schrödinger	5	1.41(205)	60%(0.53)	12.57(15)	n/a	-2.14	Prime 2.2 [30], Desmond [41, 42]	Alignment refinement, Side-chain rotamer adjustment	
Strasbourg	1	3.04(205)	60%(1.32)	7.83(15)	n/a	-2.7	ClustalW [1], GPCRmod [43], Modeller v9.1 [2]	Side chain rotamer adjustment	
Strasbourg	2	2.99(205)	64%(1.26)	8.87(15)	n/a	-2.97	ClustalW [1], GPCRmod [43], Modeller v9.1 [2]	Side chain rotamer adjustment	
LenServer	5	20.33(205)	9%(1.47)	10.82(15)	n/a	-25.67			

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
LenServer	1	20.34(205)	9%(1.47)	10.81(15)	n/a	-25.68			
LenServer	4	20.35(205)	9%(1.48)	10.81(15)	n/a	-25.68			
LenServer	3	20.47(205)	5%(1.58)	10.57(15)	n/a	-25.76			
LenServer	2	20.48(205)	5%(1.59)	10.57(15)	n/a	-25.77			
Schrödinger	2,4	1.44(205)	62%(0.54)	3.57(6)	n/a	ND	Prime 2.2 [30], Desmond [41, 42]	Alignment refinement, Side-chain rotamer adjustment	
Schrödinger	1,3	1.43(205)	63%(0.55)	3.62(6)	n/a	ND	Prime 2.2 [30], Desmond [41, 42]	Alignment refinement, Side-chain rotamer adjustment	
CXCR4/IT1t									
Baylor-Barth	4	2.42(201)	60%(1.19)	4.32(16)	10.,128.3	1.72	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		
Baylor-Barth	2	2.5(201)	64%(1.15)	5.4(16)	2.3,-2.4	1.36	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		
Baylor-Barth	5	2.63(201)	59%(1.16)	5.46(16)	11.2,-5.5	1.24	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		
VU-MedChem	3	2.14(204)	64%(1.14)	7.41(16)	18.4,43.9	1.16	ClustalW [1], MOE [6], AMBER [44], MODELLER [2]	TM1 and TM2 conformation adjustment	
PharmaDesign	2	2.57(204)	68%(1.12)	6.19(16)	29.8,67.	1.1	ClustalW [1], MOE 2007 [6]	Alignment refinement, TM2 conformation adjustment	
VU-MedChem	5	2.21(204)	64%(1.14)	7.42(16)	22.9,43.9	1.1	ClustalW [1], MOE [6], AMBER [44], MODELLER [2]	TM1 and TM2 conformation adjustment	
PharmaDesign	1	2.62(204)	66%(1.08)	6.03(16)	35.8,82.5	1.1	ClustalW [1], MOE 2007 [6]	Alignment refinement, TM2 conformation adjustment	
Monash-Sexton-1	2	2.26(204)	64%(1.08)	7.4(16)	7.1,-5.5	1.06			
PharmaDesign	3	2.67(204)	66%(1.1)	6.32(16)	33.4,63.	0.98	ClustalW [1], MOE 2007 [6]	Alignment refinement, TM2 conformation adjustment	
UMich-Zhang	3	2.27(204)	59%(1.13)	7.93(16)	70.3,60.2	0.91	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
UMich-Pogozheva	1	2.05(204)	67%(1.17)	8.78(16)	0.8,-18.7	0.89	[20, 21], DIANA [22]	Alignment refinement	
UMich-Zhang	2	2.08(204)	66%(1.14)	8.68(16)	47.9,48.3	0.88	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
UMich-Zhang	5	2.18(204)	61%(1.09)	8.35(16)	56.7,57.8	0.88	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
UMich-Zhang	1	2.28(204)	58%(1.12)	8.05(16)	69.2,59.	0.87	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
PharmaDesign	4	2.65(204)	65%(1.08)	7.05(16)	30.4,85.	0.8	ClustalW [1], MOE 2007 [6]	Alignment refinement, TM2 conformation adjustment	
PharmaDesign	5	2.69(204)	66%(1.09)	7.06(16)	36.7,28.3	0.77	ClustalW [1], MOE 2007 [6]	Alignment refinement, TM2 conformation adjustment	
Monash-Sexton-1	5	2.18(204)	66%(1.09)	8.83(16)	4.7,-3.3	0.76			
UMich-Zhang	4	2.22(204)	61%(1.12)	8.71(16)	63.,63.	0.75	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
VU-MedChem	1	2.23(204)	65%(1.14)	9.19(16)	22.6,47.6	0.62	ClustalW [1], MOE [6], AMBER [44], MODELLER [2]	TM1 and TM2 conformation adjustment	
UNC	5	2.4(204)	61%(1.07)	8.62(16)	76.9,72.8	0.62	ClustalW [1], Medusa [38], Chiron [39]		Ab-initio sampling of loops
UNC	2	2.34(204)	61%(1.07)	8.86(16)	77.,72.7	0.61	ClustalW [1], Medusa [38], Chiron [39]		Ab-initio sampling of loops
VU-MedChem	4	2.26(204)	65%(1.14)	9.14(16)	25.7,39.7	0.61	ClustalW [1], MOE [6], AMBER [44], MODELLER [2]	TM1 and TM2 conformation adjustment	
Baylor-Barth	3	2.32(201)	66%(1.15)	8.94(16)	9.,16.6	0.6	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		
UNC	4	2.37(204)	61%(1.08)	8.83(16)	76.9,72.6	0.59	ClustalW [1], Medusa [38], Chiron [39]		Ab-initio sampling of loops
VU-MedChem	2	2.29(204)	64%(1.13)	9.15(16)	23.2,47.3	0.58	ClustalW [1], MOE [6], AMBER [44], MODELLER [2]	TM1 and TM2 conformation adjustment	
UNC	1	2.39(204)	61%(1.07)	8.81(16)	76.9,72.6	0.57	ClustalW [1], Medusa [38], Chiron [39]		Ab-initio sampling of loops
Stockholm	1,2,3	2.41(204)	69%(1.11)	8.92(16)	8.1,2.	0.53	Pcons.net [27], Modeller v9.8 [2]		
Stockholm	4,5	2.46(204)	67%(1.12)	9.01(16)	8.1,2.	0.46	Pcons.net [27], Modeller v9.8 [2]		
UWash	5	2.7(204)	59%(1.15)	8.3(16)	12.7,-19.4	0.44	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
Helsinki-Xhaard	1	2.4(204)	62%(1.14)	9.41(16)	61.6,72.	0.41	Modeller v9.2 [2], SYBYL-X [24]	Side-chain rotamer adjustment	
UNC	3	2.52(204)	61%(1.07)	9.02(16)	77.1,72.9	0.41	ClustalW [1], Medusa [38], Chiron [39]		Ab-initio sampling of loops
Baylor-Barth	1	2.75(201)	52%(1.24)	8.27(16)	-77.7,-60.4	0.4	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		
UNM	5	2.87(204)	64%(1.12)	8.08(16)	88.8,82.6	0.35	3D-Jury [16], Modeller v9.8 [2]	Alignment refinement	

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
UNM	1	2.88(204)	64%(1.06)	8.07(16)	91.3,93.3	0.33	3D-Jury [16], Modeller v9.8 [2]	Alignment refinement	
UNM	3	2.88(204)	64%(1.13)	8.13(16)	96.4,116.9	0.32	3D-Jury [16], Modeller v9.8 [2]	Alignment refinement	
UNM	4	2.87(204)	64%(1.11)	8.15(16)	79.,114.4	0.32	3D-Jury [16], Modeller v9.8 [2]	Alignment refinement	
Monash-Sexton-1	1	2.22(204)	65%(1.05)	10.44(16)	-8.,-9.3	0.3			
Helsinki-Xhaard	3	2.4(204)	62%(1.1)	9.96(16)	63.9,72.5	0.27	Modeller v9.2 [2], SYBYL-X [24]	Side-chain rotamer adjustment	
Helsinki-Xhaard	2	2.38(204)	62%(1.09)	10.(16)	63.1,72.7	0.27	Modeller v9.2 [2], SYBYL-X [24]	Side-chain rotamer adjustment	
UMich-Lomize	1	2.29(204)	67%(1.18)	10.37(16)	-0.3,-15.5	0.26	[20, 21], DIANA [22]	Alignment refinement	
UNM	2	2.96(204)	64%(1.06)	8.1(16)	124.4,72.4	0.26	3D-Jury [16], Modeller v9.8 [2]	Alignment refinement	
RHUL	1,2,3,4,5	2.38(204)	61%(1.06)	10.47(16)	73.,75.4	0.15	Modeller [2]		
Sydney	2	2.66(204)	58%(1.15)	9.88(16)	73.1,57.3	0.06	Modeller v9.8 [2], MOLPROBITY [45]	Alignment refinement, Model selection	
QUB	4	2.71(204)	60%(1.02)	9.95(16)	91.6,96.3	0	ClustalW [1], MOE [6]	Alignment refinement	
QUB	5	2.71(204)	60%(1.04)	9.98(16)	91.,93.1	-0.01	ClustalW [1], MOE [6]	Alignment refinement	
QUB	2	2.7(204)	61%(1.03)	10.01(16)	91.5,94.1	-0.01	ClustalW [1], MOE [6]	Alignment refinement	
QUB	3	2.71(204)	61%(1.05)	10.08(16)	93.1,93.9	-0.04	ClustalW [1], MOE [6]	Alignment refinement	
QUB	1	2.71(204)	61%(1.05)	10.13(16)	91.5,96.7	-0.05	ClustalW [1], MOE [6]	Alignment refinement	
Soochow	1,2,3	3.36(204)	60%(1.1)	8.64(16)	6.8,-0.1	-0.23	ClustalW [1], Accelrys Discovery Studio [23], CCBTX [37]		
UWash	3	2.71(204)	60%(1.04)	10.99(16)	20.7,-26.8	-0.27	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
UWash	4	2.73(204)	59%(1.13)	11.04(16)	22.,-25.7	-0.31	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
Warsaw	3	3.21(204)	60%(1.03)	9.69(16)	87.,80.	-0.38	MUSCLE[36], ClustalW [1], Modeller [2],	Alignment refinement	
Sydney	5	2.85(204)	65%(1.14)	10.97(16)	1.2,-0.9	-0.39	Modeller v9.8 [2], MOLPROBITY [45]	Alignment refinement, Model selection	
Monash-Sexton-1	4	2.56(204)	63%(1.18)	11.95(16)	0.6,-1.4	-0.4			

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
Sydney	4	2.85(204)	65%(1.13)	10.99(16)	0.6,1.5	-0.4	Modeller v9.8 [2], MOLPROBITY [45]	Alignment refinement, Model selection	
UWash	1	2.92(204)	51%(1.19)	10.76(16)	21.3,-19.8	-0.4	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
Sydney	1	2.84(204)	65%(1.15)	11.08(16)	-1.6,0.4	-0.42	Modeller v9.8 [2], MOLPROBITY [45]	Alignment refinement, Model selection	
Sydney	3	2.86(204)	65%(1.15)	11.04(16)	6.,-4.4	-0.42	Modeller v9.8 [2], MOLPROBITY [45]	Alignment refinement, Model selection	
Monash-Sexton-1	3	2.36(204)	66%(1.17)	13.3(16)	-2.4,9.3	-0.57			
Warsaw	4,5	3.38(204)	59%(1.02)	9.9(16)	88.9,80.8	-0.58	MUSCLE[36], ClustalW [1], Modeller [2],	Alignment refinement	
Warsaw	1,2	3.24(204)	59%(1.06)	10.35(16)	90.3,78.3	-0.58	MUSCLE[36], ClustalW [1], Modeller [2],	Alignment refinement	
UWash	2	2.78(204)	60%(1.06)	12.19(16)	18.6,-29.4	-0.65	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
GaTech	1,2,3,4,5	2.98(204)	57%(1.3)	12.05(16)	3.7,7.4	-0.79	TASSERmfret [29]		
Caltech	3	3.04(204)	41%(1.26)	12.42(16)	6.9,-4.7	-0.94	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
Caltech	2	3.44(204)	35%(1.33)	11.19(16)	-2.9,13.8	-0.97	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
Evotec	4	2.94(204)	55%(1.2)	13.18(16)	11.9,0.3	-1.05	MOE [6]	TM2 conformation adjustment	
Evotec	5	2.94(204)	55%(1.2)	13.2(16)	11.9,0.3	-1.05	MOE [6]	TM2 conformation adjustment	
COH-Vaidehi	2	4.15(204)	44%(1.08)	9.19(16)	-0.2,2.	-1.08	Modeller [2], LiTiCon [28]	ECL2 and ECL3 conformation adjustment	
COH-Vaidehi	1,5	4.16(204)	44%(1.09)	9.16(16)	0.,-1.4	-1.08	Modeller [2], LiTiCon [28]	ECL2 and ECL3 conformation adjustment	
Evotec	1	2.95(204)	56%(1.19)	13.34(16)	11.9,0.4	-1.1	MOE [6]	TM2 conformation adjustment	
Evotec	2	2.94(204)	55%(1.18)	13.79(16)	11.8,2.4	-1.21	MOE [6]	TM2 conformation adjustment	
Evotec	3	2.95(204)	56%(1.2)	13.79(16)	11.8,2.6	-1.21	MOE [6]	TM2 conformation adjustment	
Caltech	1	3.63(204)	29%(1.41)	11.72(16)	14.4,2.7	-1.28	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
Caltech	4	3.69(204)	34%(1.38)	12.41(16)	20.1,-3.8	-1.5	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
Caltech	5	3.72(204)	33%(1.39)	12.93(16)	18.4,-6.	-1.66	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
COH-Vaidehi	4	5.28(204)	19%(1.57)	10.38(16)	91.5,-0.8	-2.37	MembStruk [40], Modeller [2]	ECL2 and ECL3 conformation adjustment	Ab-initio modeling of TM bundle
COH-Vaidehi	3	5.28(204)	18%(1.56)	10.42(16)	92.,0.1	-2.39	MembStruk [40], Modeller [2]	ECL2 and ECL3 conformation adjustment	Ab-initio modeling of TM bundle
LenServer	2	11.19(204)	1%(1.66)	24.28(16)	166.,177.1	-11.2			
LenServer	5	11.21(204)	1%(1.62)	24.21(16)	166.1,177.3	-11.2			
LenServer	1	11.22(204)	1%(1.7)	24.18(16)	165.9,177.7	-11.2			
LenServer	3	11.21(204)	1%(1.61)	24.23(16)	165.8,176.8	-11.21			
LenServer	4	11.23(204)	1%(1.86)	24.61(16)	167.3,173.2	-11.32			
UCSF-Shoichet-2	4	2.53(203)	58%(1.15)	0.(0)	2.4,-19.8	ND	PROMALS-3D [17], Modeller v9.8 [2], 3K-ENM [18], PLOP [46]	Alignment refinement	
UCSF-Shoichet-2	3	2.75(203)	63%(1.17)	0.(0)	2.8,1.2	ND	PROMALS-3D [17], Modeller v9.8 [2], 3K-ENM [18], PLOP [46]	Alignment refinement	
UCSF-Shoichet-2	5	2.79(204)	67%(1.15)	0.(0)	31.5,-0.4	ND	PROMALS-3D [17], Modeller v9.8 [2], 3K-ENM [18], PLOP [46]	Alignment refinement	
UCSF-Shoichet-2	2	2.81(204)	64%(1.14)	0.(0)	11.2,-0.6	ND	PROMALS-3D [17], Modeller v9.8 [2], 3K-ENM [18], PLOP [46]	Alignment refinement	
UCSF-Shoichet-2	1	2.83(203)	62%(1.13)	0.(0)	26.6,-1.7	ND	PROMALS-3D [17], Modeller v9.8 [2], 3K-ENM [18], PLOP [46]	Alignment refinement	
CDD-CMBI	2	3.14(204)	58%(0.97)	7.59(11)	78.,75.9	ND	Yasara[35]	Model refinement	
CDD-CMBI	5	3.14(204)	57%(0.95)	7.59(11)	83.,65.2	ND	Yasara[35]	Model refinement	
CDD-CMBI	3	3.14(204)	57%(0.96)	7.61(11)	85.9,72.1	ND	Yasara[35]	Model refinement	
CDD-CMBI	4	3.15(204)	57%(0.96)	7.6(11)	82.3,76.2	ND	Yasara[35]	Model refinement	
CDD-CMBI	1	3.15(204)	57%(0.95)	7.61(11)	79.2,75.9	ND	Yasara[35]	Model refinement	
CXCR4/CVX15									
UMich-Zhang	2	2.82(204)	60%(1.19)	6.64(16)	74.,60.	1.59	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
UMich-Zhang	1	2.82(204)	60%(1.18)	6.78(16)	73.1,67.3	1.54	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
Baylor-Barth	1	2.87(201)	63%(1.09)	7.07(16)	-20.6,-25.4	1.41	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		No ligand
PharmaDesign	1	3.08(204)	65%(1.03)	6.61(16)	30.4,74.4	1.4	ClustalW [1], MOE 2007 [6]	Alignment refinement, CXCR4 TM2 conformation adjustment	
UMich-Zhang	3	2.86(204)	59%(1.15)	7.91(16)	74.4,59.6	1.15	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
Baylor-Barth	2	2.79(201)	62%(1.12)	8.07(16)	-5.4,1.9	1.15	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		No ligand
UMich-Zhang	5	2.88(204)	53%(1.08)	8.19(16)	58.8,57.5	1.04	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
Baylor-Barth	4	2.76(201)	62%(1.14)	8.67(16)	-8.5,12.6	0.98	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		No ligand
UMich-Pogozheva	1	2.53(204)	64%(1.)	9.53(16)	-0.8,-20.1	0.88	[20, 21], DIANA [22]	Alignment refinement	
Stockholm	1,2,5	2.97(204)	65%(1.06)	8.97(16)	10.5,4.9	0.72	Pcons.net [27], Modeller v9.8 [2]		
UNC	2	2.79(204)	64%(1.1)	9.62(16)	78.5,74.3	0.64	ClustalW [1], Medusa [38], Chiron [39]		Ab-initio sampling of loops
UNC	5	2.8(204)	64%(1.1)	9.7(16)	78.5,74.3	0.62	ClustalW [1], Medusa [38], Chiron [39]		Ab-initio sampling of loops
UNC	1	2.81(204)	64%(1.1)	9.7(16)	78.5,74.1	0.61	ClustalW [1], Medusa [38], Chiron [39]		Ab-initio sampling of loops
UMich-Zhang	4	3.22(204)	51%(1.01)	8.73(16)	66.,65.1	0.6	LOMETS [12], I-TASSER [14], SPICKER [13], FG-MD [15]		
UNC	3	2.83(204)	64%(1.09)	9.7(16)	78.5,74.1	0.59	ClustalW [1], Medusa [38], Chiron [39]		Ab-initio sampling of loops
UNC	4	2.84(204)	64%(1.1)	9.76(16)	78.5,74.1	0.57	ClustalW [1], Medusa [38], Chiron [39]		Ab-initio sampling of loops
UWash	5	3.22(204)	62%(1.13)	9.27(16)	11.1,-22.4	0.43	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
Baylor-Barth	5	2.78(201)	64%(1.1)	10.42(16)	-9.1,-2.7	0.4	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		No ligand
Helsinki-Xhaard	1	2.82(204)	62%(1.07)	10.5(16)	66.6,75.5	0.34	Modeller v9.2 [2], SYBYL-X [24]	Side-chain rotamer adjustment	
UMich-Lomize	1	2.83(204)	65%(1.12)	10.59(16)	-2.6,-13.8	0.31	[20, 21], DIANA [22]	Alignment refinement	
MolLife	1	3.68(149)	37%(1.44)	8.89(16)	53.8,76.	0.2	Enhanced homology modeling [47, 48]		
RHUL	1,2,3,4,5	2.84(204)	60%(1.05)	10.89(16)	75.4,77.	0.2	Modeller [2]		

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
QUB	2	3.07(204)	63%(1.03)	10.36(16)	96.3,96.8	0.2	ClustalW [1], MOE [6]	Alignment refinement	
QUB	1	3.08(204)	63%(1.03)	10.37(16)	95.6,98.4	0.19	ClustalW [1], MOE [6]	Alignment refinement	
Baylor-Barth	3	3.35(201)	56%(1.14)	10.04(16)	34.6,48.6	0.09	Comparative modeling technique developed for membrane protein structures (Chen et al, in preparation), Rosetta-Membrane [26]		No ligand
MolLife	5	4.09(149)	27%(1.37)	8.46(16)	63.,84.8	0.03	Enhanced homology modeling [47, 48]		
Soochow	1,2,3,4,5	3.79(204)	57%(1.04)	9.53(16)	4.8,-1.7	-0.08	ClustalW [1], Accelrys Discovery Studio [23], CCBTX [37]		
MolLife	4	3.8(149)	33%(1.5)	9.65(16)	60.4,86.5	-0.13	Enhanced homology modeling [47, 48]		
Stockholm	3,4	2.97(204)	65%(1.06)	11.65(16)	10.5,4.9	-0.15	Pcons.net [27], Modeller v9.8 [2]		
UWash	3	3.17(204)	61%(1.07)	11.58(16)	19.7,-29.4	-0.28	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
MolLife	2	3.88(149)	32%(1.56)	9.96(16)	60.6,92.3	-0.29	Enhanced homology modeling [47, 48]		
UWash	4	3.08(204)	61%(1.13)	11.86(16)	20.5,-28.1	-0.29	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
UWash	1	3.28(204)	51%(1.14)	11.54(16)	19.3,-22.4	-0.35	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
Evotec	3	3.62(204)	56%(1.2)	10.87(16)	14.6,2.8	-0.39	MOE [6]	TM2 conformation adjustment	
Evotec	5	3.53(204)	56%(1.2)	11.1(16)	14.8,3.3	-0.4	MOE [6]	TM2 conformation adjustment	
Evotec	4	3.58(204)	56%(1.2)	11.06(16)	14.8,3.3	-0.42	MOE [6]	TM2 conformation adjustment	
Evotec	2	3.58(204)	56%(1.2)	11.83(16)	14.8,3.3	-0.67	MOE [6]	TM2 conformation adjustment	
Evotec	1	3.54(204)	56%(1.2)	12.03(16)	14.8,3.3	-0.7	MOE [6]	TM2 conformation adjustment	
UWash	2	3.15(204)	62%(1.07)	13.19(16)	17.7,-32.	-0.78	HHPred [31], Rosetta-Membrane [32]		Coarse-grained membrane environment, No ligand
GaTech	3	3.46(204)	58%(1.21)	12.55(16)	1.9,16.3	-0.81	TASSERmfret [29]		
MolLife	3	4.15(149)	29%(1.53)	11.02(16)	51.2,83.7	-0.84	Enhanced homology modeling [47, 48]		
GaTech	2	3.48(204)	58%(1.23)	12.65(16)	1.8,14.1	-0.85	TASSERmfret [29]		

Group	Model	TM RMSD, A (# residues)	fraction TM superimposed (RMSD)	ECL2 RMSD, A (# of residues)	W94 and D97 rotation, degrees	Protein prediction z-score	Software	Manual steps	Comments
GaTech	4	3.48(204)	58%(1.2)	12.67(16)	13.2,14.2	-0.86	TASSERmfret [29]		
GaTech	1	3.47(204)	57%(1.22)	12.73(16)	3.6,21.	-0.87	TASSERmfret [29]		
COH-Vaidehi	1,3,4,5	4.99(204)	49%(1.03)	9.3(16)	0.,4.4	-0.93	Modeller [2], LiTiCon [28]	ECL2 and ECL3 conformation adjustment	
Caltech	5	4.28(204)	22%(1.47)	11.49(16)	59.3,21.1	-1.1	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
Caltech	3	5.01(204)	22%(1.39)	10.17(16)	49.3,11.9	-1.22	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
Caltech	1	4.65(204)	23%(1.38)	13.24(16)	25.4,38.5	-1.94	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
Caltech	2	4.72(204)	28%(1.49)	13.44(16)	28.9,51.9	-2.05	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
Caltech	4	5.(204)	19%(1.49)	13.65(16)	50.4,39.4	-2.34	BiHelix, CombiHelix, SuperBihelix, SuperCombiHelix [33] , MacroModel [34]	Alignment refinement	Ab-initio modeling of TM bundle
COH-Vaidehi	2	5.72(204)	23%(1.36)	11.99(16)	56.6,-3.8	-2.35	MembStruk [40], Modeller [2]	ECL2 and ECL3 conformation adjustment	Ab-initio modeling of TM bundle

Table S2. Ligand complex prediction accuracy in GPCR Dock 2010, related to Figures 3, 5A, 6, 7 and Table 1.

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
D3/eticlopride											
PompeuFabra	3	1.5(15)	1.16(14)	63%	0.96	36(9)/65(15)	39.95(58%)	2.39	GOLD [49], MOE [6]		
CDD-CMBI	1	2.25(15)	1.67(14)	50%	2.13	36(12)/65(15)	38.75(57%)	2.05	Yasara [50], Snooker, Fleksy [51]		Ligand binding mode prediction with structure-based pharmacophores
COH-Vaidehi	1	2.04(15)	1.36(14)	59%	1.22	31(10)/65(15)	34.78(51%)	2	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
UCSF-Shoichet-1	4	1.53(15)	1.41(14)	74%	1.23	29(10)/65(15)	32.46(47%)	1.85	DOCK3.5.54 [54]	Pose selection	Model selection by VLS enrichment
Schrödinger	5	1.57(15)	1.48(14)	38%	1.77	23(11)/65(15)	30.94(45%)	1.63	IFD [55], WaterMap [56], Phase [57]	Pose generation, Pose selection	Ligand pharmacophore and WaterMap based model validation
CDD-CMBI	2	2.24(15)	1.67(14)	46%	2.27	26(11)/65(15)	30.5(44%)	1.49	Yasara [50], Snooker, Fleksy [51]		Ligand binding mode prediction with structure-based pharmacophores
CDD-CMBI	3	2.17(15)	1.59(14)	43%	2.38	30(11)/65(15)	30.1(44%)	1.44	Yasara [50], Snooker, Fleksy [51]		Ligand binding mode prediction with structure-based pharmacophores
Warsaw	4	1.68(15)	1.64(14)	52%	2.34	27(9)/65(15)	26.47(39%)	1.22	Glide 5.6 [52], AutoDock [58]	Pose selection	
CDD-CMBI	4	1.99(15)	1.62(14)	35%	2.13	19(8)/65(15)	24.84(36%)	1.16	Yasara [50], Snooker, Fleksy [51]		Ligand binding mode prediction with structure-based pharmacophores
CDD-CMBI	5	1.87(15)	1.67(14)	41%	2.44	21(8)/65(15)	25.57(37%)	1.14	Yasara [50], Snooker, Fleksy [51]		Ligand binding mode prediction with structure-based pharmacophores
Helsinki-Xhaard	1	1.43(15)	1.29(14)	47%	3.42	21(7)/65(15)	28.95(42%)	1.13	SYBYL-X [24]	Small molecule docking	Manual docking
COH-Vaidehi	2	2.16(15)	1.5(14)	57%	2.96	25(7)/65(15)	26.75(39%)	1.1	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
Monash-Sexton-2	3	2.47(15)	2.3(14)	52%	1.94	17(6)/65(15)	22.51(33%)	1.05	Surflex [59]		Model selection by VLS enrichment
Schrödinger	1	1.5(15)	1.42(14)	61%	2.24	19(10)/65(15)	22.81(33%)	1.01	IFD [55], WaterMap [56], Phase [57]	Pose generation, Pose selection	Ligand pharmacophore and WaterMap based model validation
Monash-Hall	3	2.4(15)	2.05(14)	47%	2.1	18(10)/65(15)	22.05(32%)	0.99	ICM [60]		Model selection by VLS enrichment
UNC	5	1.77(15)	1.7(14)	58%	3.25	20(6)/65(15)	24.28(35%)	0.87	MedusaDock [38], DMD [61]	Pose selection	
COH-Vaidehi	3	2.03(15)	1.43(14)	59%	2.5	18(8)/65(15)	21.43(31%)	0.86	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
Monash-Yuriev	3	2.38(15)	1.72(14)	55%	2.33	20(8)/65(15)	20.81(30%)	0.86	Glide 5.6 [52], IFD [55]	Pose selection	Model selection by VLS enrichment
Monash-Hall	1	2.18(15)	2.04(14)	35%	2.21	16(11)/65(15)	19.68(29%)	0.81	ICM [60]		Model selection by VLS enrichment
Schrödinger	2	1.87(15)	1.85(14)	61%	2.68	17(8)/65(15)	20.08(29%)	0.73	IFD [55], WaterMap [56], Phase [57]	Pose generation, Pose selection	Ligand pharmacophore and WaterMap based model validation
UCSF-Shoichet-1	1	2.17(15)	1.96(14)	68%	2.39	17(6)/65(15)	18.79(27%)	0.71	DOCK3.5.54 [54]	Pose selection	Model selection by VLS enrichment
UMich-Zhang	1	2.01(15)	1.93(14)	53%	3.42	17(5)/65(15)	22.32(33%)	0.71	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
Monash-Hall	2	2.62(15)	2.08(14)	34%	2.16	13(9)/65(15)	17.85(26%)	0.7	ICM [60]		Model selection by VLS enrichment
Monash-Sexton-2	2	2.19(15)	2.06(14)	53%	2.98	16(6)/65(15)	20.1(29%)	0.67	Surflex [59]		Model selection by VLS enrichment
QUB	4	1.76(15)	1.66(14)	54%	3.29	17(6)/65(15)	21.08(31%)	0.66	Glide 5.6 [52], IFD [55], Prime [30], MacroModel [34]	Pose selection	
Soochow	1	1.69(15)	1.63(14)	45%	2.24	16(6)/65(15)	16.96(25%)	0.63	Discovery Studio [23]		
Monash-Hall	4	2.15(15)	2.11(14)	41%	2.09	9(8)/65(15)	16.17(24%)	0.61	ICM [60]		Model selection by VLS enrichment
Stockholm	2	1.54(15)	1.58(14)	53%	2.98	11(7)/65(15)	18.01(26%)	0.53	MacroModel [34], Glide 5.5 [52], AutoDock Vina [64]		
UNM	1	1.57(15)	1.46(14)	51%	3.58	17(5)/65(15)	19.29(28%)	0.48	Glide [52]	Pose selection	
UNM	2	1.57(15)	1.46(14)	47%	4.33	22(6)/65(15)	21.15(31%)	0.43	Glide [52]	Pose selection	
GaTech	2	3.05(15)	3.09(14)	56%	2.18	11(7)/65(15)	13.33(19%)	0.41	Q-Dock [65], AMBER 8 [44]		
Stockholm	3	1.54(15)	1.57(14)	54%	2.84	11(4)/65(15)	15.03(22%)	0.37	MacroModel [34], Glide 5.5 [52], AutoDock Vina [64]		
UNC	3	1.79(15)	1.73(14)	52%	3.05	12(4)/65(15)	15.58(23%)	0.36	MedusaDock [38], DMD [61]	Pose selection	
WUStL	2	1.51(15)	1.5(14)	60%	3.11	12(5)/65(15)	15.67(23%)	0.35	SYBYL 7.2 [24], GOLD 4.0 [49]		MD performed on explicit membrane + water
UMich-Zhang	2	2.31(15)	1.9(14)	53%	4.22	18(6)/65(15)	19.53(28%)	0.35	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
Evotec	1	2.76(15)	2.79(14)	43%	3.94	16(6)/65(15)	18.48(27%)	0.35	GOLD [49]	Pose selection	
Caltech	4	2.89(15)	2.79(14)	76%	3.45	11(2)/65(15)	16.52(24%)	0.33	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
UNSW	2	1.68(15)	1.59(14)	63%	3.68	14(4)/65(15)	17.1(25%)	0.32	Accelrys Discovery Studio [23], Gold 4.1 [49], PROCHECK [7]	Pose selection	Ligand binding mode prediction with structure-based pharmacophores
UNSW	1	1.64(15)	1.58(14)	58%	4.07	18(5)/65(15)	18.34(27%)	0.31	Accelrys Discovery Studio [23], Gold 4.1 [49], PROCHECK [7]	Pose selection	Ligand binding mode prediction with structure-based pharmacophores
UMich-Lomize	1	1.62(15)	1.49(14)	54%	3.26	9(5)/65(15)	14.53(21%)	0.25	QUANTA [66], Glide[52]	Ligand docking	Ligand structure errors
UMich-Pogozheva	1	1.62(15)	1.49(14)	54%	3.26	9(5)/65(15)	14.53(21%)	0.25	QUANTA [66], Glide[52]	Ligand docking	Ligand structure errors
Monash-Yuriev	5	2.21(15)	1.55(14)	49%	3.46	15(5)/65(15)	15.19(22%)	0.24	Glide 5.6 [52], IFD [55]	Pose selection	Model selection by VLS enrichment
Monash-Yuriev	1	2.36(15)	1.74(14)	51%	2.95	11(4)/65(15)	13.11(19%)	0.22	Glide 5.6 [52], IFD [55]	Pose selection	Model selection by VLS enrichment
UCSF-Shoichet-1	2	1.55(15)	1.47(14)	60%	3.3	14(5)/65(15)	14.24(21%)	0.22	DOCK3.5.54 [54]	Pose selection	Model selection by VLS enrichment
Stockholm	1	1.54(15)	1.58(14)	53%	3.11	7(3)/65(15)	13.53(20%)	0.22	MacroModel [34], Glide 5.5 [52], AutoDock Vina [64]		
GaTech	1	3.05(15)	3.09(14)	54%	2.62	9(4)/65(15)	11.43(17%)	0.19	Q-Dock [65], AMBER 8 [44]		
UNC	2	1.81(15)	1.76(14)	40%	3.84	13(5)/65(15)	15.65(23%)	0.19	MedusaDock [38], DMD [61]	Pose selection	
WUStL	1	1.65(15)	1.43(14)	71%	3.85	15(3)/65(15)	15.42(22%)	0.17	SYBYL 7.2 [24], GOLD 4.0 [49]		MD performed on explicit membrane + water
UCSF-Shoichet-1	5	1.75(15)	1.63(14)	62%	3.41	11(4)/65(15)	13.77(20%)	0.16	DOCK3.5.54 [54]	Pose selection	Model selection by VLS enrichment
QUB	1	1.93(15)	1.74(14)	54%	4.19	12(4)/65(15)	16.3(24%)	0.15	Glide 5.6 [52], IFD [55], Prime [30], MacroModel [34]	Pose selection	
UNM	3	1.59(15)	1.48(14)	53%	4.93	18(6)/65(15)	18.82(27%)	0.15	Glide [52]	Pose selection	
Monash-Yuriev	2	2.26(15)	1.53(14)	45%	3.44	11(5)/65(15)	13.34(19%)	0.13	Glide 5.6 [52], IFD [55]	Pose selection	Model selection by VLS enrichment
UMich-Zhang	3	3.1(15)	1.99(14)	70%	4.47	12(5)/65(15)	16.64(24%)	0.11	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
PompeuFabra	2	1.38(15)	1.24(14)	61%	3.4	11(3)/65(15)	12.88(19%)	0.11	GOLD [49], MOE [6]		
QUB	5	1.91(15)	1.8(14)	44%	3.81	10(4)/65(15)	14.24(21%)	0.11	Glide 5.6 [52], IFD [55], Prime [30], MacroModel [34]	Pose selection	
Evotec	5	2.76(15)	2.79(14)	47%	3.91	14(6)/65(15)	14.35(21%)	0.09	GOLD [49]	Pose selection	
Schrödinger	4	1.87(15)	1.85(14)	61%	3.53	11(4)/65(15)	13.(19%)	0.09	IFD [55], WaterMap [56], Phase [57]	Pose generation, Pose selection	Ligand pharmacophore and WaterMap based model validation

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
GaTech	3	3.05(15)	3.09(14)	57%	2.41	7(4)/65(15)	8.85(13%)	0.07	Q-Dock [65], AMBER 8 [44]		
PompeuFabra	1	1.45(15)	1.33(14)	50%	3.62	10(4)/65(15)	12.44(18%)	0.03	GOLD [49], MOE [6]		
Evotec	3	2.76(15)	2.79(14)	44%	3.84	12(5)/65(15)	12.97(19%)	0.02	GOLD [49]	Pose selection	
Caltech	2	2.64(15)	2.12(14)	81%	4.13	10(2)/65(15)	13.54(20%)	-0.01	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
WUStL	3	1.57(15)	1.57(14)	74%	4.55	15(6)/65(15)	14.76(22%)	-0.03	SYBYL 7.2 [24], GOLD 4.0 [49]		MD performed on explicit membrane + water
Monash-Sexton-2	5	2.48(15)	2.19(14)	54%	3.89	10(7)/65(15)	12.11(18%)	-0.05	Surflex [59]		Model selection by VLS enrichment
UMich-Zhang	4	2.51(15)	1.87(14)	65%	4.28	12(4)/65(15)	13.07(19%)	-0.07	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
KIAS	5	2.17(15)	1.59(14)	47%	3.51	6(5)/65(15)	10.37(15%)	-0.08	Glide [52]	Pose selection	
Schrödinger	3	1.5(15)	1.42(14)	61%	3.71	10(3)/65(15)	10.91(16%)	-0.09	IFD [55], WaterMap [56], Phase [57]	Pose generation, Pose selection	Ligand pharmacophore and WaterMap based model validation
QUB	3	1.86(15)	1.64(14)	65%	5.24	12(5)/65(15)	15.8(23%)	-0.11	Glide 5.6 [52], IFD [55], Prime [30], MacroModel [34]	Pose selection	
KIAS	1	2.83(15)	2.32(14)	43%	3.94	6(3)/65(15)	11.02(16%)	-0.13	Glide [52]	Pose selection	
COH-Vaidehi	5	2.79(15)	2.57(14)	48%	4.09	10(4)/65(15)	11.01(16%)	-0.16	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
Monash-Hall	5	2.16(15)	1.44(14)	64%	3.64	5(4)/65(15)	9.37(14%)	-0.17	ICM [60]		Model selection by VLS enrichment
Evotec	2	2.76(15)	2.79(14)	45%	4.6	11(4)/65(15)	12.43(18%)	-0.19	GOLD [49]	Pose selection	
Caltech	3	2.88(15)	2.91(14)	60%	4.25	10(2)/65(15)	11.02(16%)	-0.2	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
Caltech	1	3.07(15)	2.43(14)	74%	5.68	13(3)/65(15)	15.47(23%)	-0.23	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
RHUL	2	1.69(15)	1.18(14)	43%	3.85	6(4)/65(15)	8.87(13%)	-0.25	LigDoc [67]		
Monash-Yuriev	4	2.31(15)	1.67(14)	46%	3.98	8(3)/65(15)	8.88(13%)	-0.27	Glide 5.6 [52], IFD [55]	Pose selection	Model selection by VLS enrichment
GaTech	5	3.05(15)	3.09(14)	52%	3.65	5(3)/65(15)	7.75(11%)	-0.28	Q-Dock [65], AMBER 8 [44]		

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
PharmaDesign	1	2.14(15)	1.78(14)	41%	4.83	11(3)/65(15)	10.99(16%)	-0.33	MOE Dock [6]	Model selection, Side-chain and loop modifications	3D pharmacophore search based on experimental anchors
Evotec	4	2.75(15)	2.78(14)	45%	4.29	7(2)/65(15)	8.73(13%)	-0.36	GOLD [49]	Pose selection	
COH-Vaidehi	4	2.77(15)	2.54(14)	56%	5.69	13(3)/65(15)	13.33(19%)	-0.37	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
Monash-Sexton-2	4	2.08(15)	1.87(14)	42%	4.28	8(3)/65(15)	7.62(11%)	-0.42	Surflex [59]		Model selection by VLS enrichment
GaTech	4	3.05(15)	3.09(14)	60%	3.42	2(1)/65(15)	3.84(6%)	-0.47	Q-Dock [65], AMBER 8 [44]		
QUB	2	1.89(15)	1.72(14)	48%	5.86	10(3)/65(15)	12.01(18%)	-0.49	Glide 5.6 [52], IFD [55], Prime [30], MacroModel [34]	Pose selection	
Monash-Sexton-2	1	2.27(15)	1.98(14)	47%	5.19	9(2)/65(15)	9.57(14%)	-0.5	Surflex [59]		Model selection by VLS enrichment
UMich-Zhang	5	2.76(15)	2.26(14)	74%	7.18	13(3)/65(15)	13.67(20%)	-0.68	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
KIAS	3	2.72(15)	2.17(14)	48%	5.88	9(5)/65(15)	9.07(13%)	-0.69	Glide [52]	Pose selection	
UNM	5	1.49(15)	1.41(14)	38%	6.99	7(4)/65(15)	11.57(17%)	-0.77	Glide [52]	Pose selection	
UNSW	5	1.74(15)	1.51(14)	40%	6.83	8(2)/65(15)	9.72(14%)	-0.86	Accelrys Discovery Studio [23], Gold 4.1 [49], PROCHECK [7]	Pose selection	Ligand binding mode prediction with structure-based pharmacophores
Baylor-Barth	3	2.13(15)	1.81(14)	36%	5.77	3(1)/65(15)	5.44(8%)	-0.9	Rosetta [4], Omega [68]	Pose selection	
UNC	1	1.79(15)	1.72(14)	43%	6.53	5(3)/65(15)	7.46(11%)	-0.93	MedusaDock [38], DMD [61]	Pose selection	
UCSF-Shoichet-1	3	2.2(15)	1.94(14)	73%	8.01	10(1)/65(15)	11.68(17%)	-0.99	DOCK3.5.54 [54]	Pose selection	Model selection by VLS enrichment
UNC	4	1.76(15)	1.69(14)	60%	6.6	4(1)/65(15)	5.86(9%)	-1.05	MedusaDock [38], DMD [61]	Pose selection	
Baylor-Barth	2	2.1(15)	1.66(14)	30%	5.7	2(2)/65(15)	2.59(4%)	-1.06	Rosetta [4], Omega [68]	Pose selection	
Stockholm	4	1.91(15)	1.57(14)	35%	5.74	1(1)/65(15)	2.48(4%)	-1.08	MacroModel [34], Glide 5.5 [52], AutoDock Vina [64]		
Caltech	5	2.98(15)	2.87(14)	62%	7.23	8(2)/65(15)	7.65(11%)	-1.08	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
RHUL	1	1.69(15)	1.18(14)	45%	6.75	5(3)/65(15)	5.69(8%)	-1.1	LigDoc [67]		

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
Stockholm	5	1.61(15)	1.58(14)	57%	7.45	5(2)/65(15)	6.69(10%)	-1.19	MacroModel [34], Glide 5.5 [52], AutoDock Vina [64]		
Warsaw	5	1.68(15)	1.64(14)	55%	7.65	8(2)/65(15)	7.08(10%)	-1.21	Glide 5.6 [52], AutoDock [58]	Pose selection	
Warsaw	1	1.68(15)	1.64(14)	52%	7.66	7(3)/65(15)	6.99(10%)	-1.22	Glide 5.6 [52], AutoDock [58]	Pose selection	
UNSW	4	1.65(15)	1.59(14)	48%	7.99	8(1)/65(15)	7.75(11%)	-1.24	Accelrys Discovery Studio [23], Gold 4.1 [49], PROCHECK [7]	Pose selection	Ligand binding mode prediction with structure-based pharmacophores
UNM	4	1.61(15)	1.5(14)	60%	8.79	9(3)/65(15)	10.41(15%)	-1.25	Glide [52]	Pose selection	
Warsaw	2	1.68(15)	1.64(14)	52%	7.78	7(3)/65(15)	6.56(10%)	-1.27	Glide 5.6 [52], AutoDock [58]	Pose selection	
UNSW	3	1.62(15)	1.56(14)	51%	8.63	9(4)/65(15)	9.36(14%)	-1.28	Accelrys Discovery Studio [23], Gold 4.1 [49], PROCHECK [7]	Pose selection	Ligand binding mode prediction with structure-based pharmacophores
Warsaw	3	1.68(15)	1.64(14)	52%	7.84	5(3)/65(15)	6.44(9%)	-1.29	Glide 5.6 [52], AutoDock [58]	Pose selection	
Baylor-Barth	1	2.17(15)	1.92(14)	41%	7.5	3(2)/65(15)	3.24(5%)	-1.42	Rosetta [4], Omega [68]	Pose selection	
KIAS	2	2.85(15)	2.39(14)	39%	7.47	2(2)/65(15)	2.32(3%)	-1.47	Glide [52]	Pose selection	
KIAS	4	2.84(15)	2.49(14)	47%	7.64	3(2)/65(15)	2.66(4%)	-1.49	Glide [52]	Pose selection	
Baylor-Barth	4	2.52(15)	2.33(14)	46%	7.96	2(1)/65(15)	2.35(3%)	-1.58	Rosetta [4], Omega [68]	Pose selection	
Strasbourg	1	2.42(15)	2.33(14)	72%	10.08	8(2)/65(15)	9.08(13%)	-1.62	SYBYL-X [24], GOLD 3.1 [49], AMBER 8 [44]	Pose selection	MD performed on explicit membrane + water
Strasbourg	2	2.67(15)	2.49(14)	69%	9.94	8(2)/65(15)	7.74(11%)	-1.68	SYBYL-X [24], GOLD 3.1 [49], AMBER 8 [44]	Pose selection	MD performed on explicit membrane + water
RHUL	4	1.69(15)	1.18(14)	12%	14.56	0(0)/65(15)	0.(0%)	-3.2	LigDoc [67]		
RHUL	5	1.69(15)	1.18(14)	2%	15.38	0(0)/65(15)	0.2(0%)	-3.37	LigDoc [67]		
RHUL	3	1.69(15)	1.18(14)	2%	15.6	0(0)/65(15)	0.(0%)	-3.43	LigDoc [67]		
LenServer	2	21.39(15)	21.97(14)	0%	17.69	0(0)/65(15)	0.(0%)	-3.9			
LenServer	3	21.26(15)	21.85(14)	0%	19.88	0(0)/65(15)	0.(0%)	-4.39			
LenServer	5	21.46(15)	22.(14)	0%	21.23	0(0)/65(15)	0.(0%)	-4.69			

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
LenServer	4	21.48(15)	22.02(14)	0%	21.37	0(0)/65(15)	0.(0%)	-4.72			
LenServer	1	21.4(15)	21.94(14)	0%	21.94	0(0)/65(15)	0.(0%)	-4.84			
CXCR4/IT1t											
VU-MedChem	5	4.94(13)	3.04(7)	47%	4.88	19(5)/64(13)	21.16(36%)	3.49	GOLD [49], AMBER [44, 69]	Pose selection	
COH-Vaidehi	1	6.01(13)	3.2(7)	47%	2.47	9(2)/64(13)	10.56(18%)	2.19	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
COH-Vaidehi	2	5.87(13)	3.07(7)	43%	2.88	8(4)/64(13)	9.93(17%)	2.02	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
COH-Vaidehi	3	7.58(13)	6.(7)	36%	6.86	10(2)/64(13)	10.28(17%)	1.47	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
UCSF-Shoichet-2	5	2.15(8)	2.18(7)	40%	6.14	7(4)/64(13)	9.22(15%)	1.41	DOCK3.5.54 [54], PLOP [46]	Pose selection	Model selection by VLS enrichment
COH-Vaidehi	5	5.98(13)	3.21(7)	38%	5.17	7(4)/64(13)	7.7(13%)	1.32	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
GaTech	3	8.05(13)	1.9(7)	30%	7.49	6(2)/64(13)	8.58(14%)	1.1	Q-Dock [65], AMBER 8 [44]		
GaTech	2	8.05(13)	1.9(7)	29%	6.41	5(2)/64(13)	7.41(12%)	1.08	Q-Dock [65], AMBER 8 [44]		
Caltech	2	6.93(13)	3.47(7)	30%	9.46	9(2)/64(13)	9.91(17%)	1.01	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
GaTech	5	8.05(13)	1.9(7)	34%	4.2	3(1)/64(13)	4.45(7%)	0.96	Q-Dock [65], AMBER 8 [44]		
Warsaw	5	5.25(13)	4.01(7)	29%	5.35	4(2)/64(13)	4.82(8%)	0.84	Glide 5.6 [52], AutoDock [58]	Pose selection	
UNM	1	4.67(13)	4.71(7)	37%	5.86	4(3)/64(13)	5.09(9%)	0.8	Glide [52]	Pose selection	
PharmaDesign	4	4.81(13)	4.65(7)	49%	6.06	5(3)/64(13)	4.88(8%)	0.74	MOE Dock [6]	Model selection, Side-chain and loop modifications	3D pharmacophore search based on experimental anchors
CDD-CMBI	2	5.08(12)	3.9(7)	20%	6.49	4(2)/64(13)	5.27(9%)	0.74	Yasara [50], Snooker, Fleksy [51]		Ligand binding mode prediction with structure-based pharmacophores
Stockholm	3	5.21(13)	2.34(7)	36%	6.23	4(2)/64(13)	4.96(8%)	0.73	MacroModel [34], Glide 5.5 [52], AutoDock Vina [64]		
VU-MedChem	2	5.(13)	3.47(7)	34%	5.86	3(1)/64(13)	4.49(8%)	0.71	GOLD [49], AMBER [44, 69]	Pose selection	

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
CDD-CMBI	5	4.94(12)	3.95(7)	27%	6.06	5(3)/64(13)	4.59(8%)	0.69	Yasara [50], Snooker, Fleksy [51]		Ligand binding mode prediction with structure-based pharmacophores
Soochow	3	4.9(13)	1.96(7)	33%	7.3	5(3)/64(13)	5.72(10%)	0.68	Discovery Studio [23]		
Baylor-Barth	1	6.06(13)	3.83(7)	27%	7.03	5(2)/64(13)	4.56(8%)	0.54	Rosetta [4], Omega [68]	Pose selection	
UCSF-Shoichet-2	4	3.34(8)	3.41(7)	42%	5.29	1(1)/64(13)	2.7(5%)	0.51	DOCK3.5.54 [54], PLOP [46]	Pose selection	Model selection by VLS enrichment
Stockholm	2	5.21(13)	2.34(7)	37%	7.15	3(2)/64(13)	4.46(7%)	0.51	MacroModel [34], Glide 5.5 [52], AutoDock Vina [64]		
Warsaw	4	5.25(13)	4.01(7)	22%	9.63	6(1)/64(13)	6.83(11%)	0.5	Glide 5.6 [52], AutoDock [58]	Pose selection	
GaTech	1	8.05(13)	1.9(7)	36%	5.41	1(1)/64(13)	2.69(5%)	0.49	Q-Dock [65], AMBER 8 [44]		
UCSF-Shoichet-2	3	2.34(8)	2.4(7)	27%	7.12	5(1)/64(13)	4.29(7%)	0.49	DOCK3.5.54 [54], PLOP [46]	Pose selection	Model selection by VLS enrichment
VU-MedChem	1	5.21(13)	3.32(7)	20%	10.08	6(2)/64(13)	7.14(12%)	0.48	GOLD [49], AMBER [44, 69]	Pose selection	
UCSF-Shoichet-2	1	2.21(8)	2.23(7)	40%	7.95	4(1)/64(13)	4.81(8%)	0.44	DOCK3.5.54 [54], PLOP [46]	Pose selection	Model selection by VLS enrichment
Caltech	1	5.8(13)	3.28(7)	21%	9.49	6(1)/64(13)	6.23(10%)	0.43	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
CDD-CMBI	4	4.82(12)	4.16(7)	25%	7.63	4(2)/64(13)	4.3(7%)	0.41	Yasara [50], Snooker, Fleksy [51]		Ligand binding mode prediction with structure-based pharmacophores
CDD-CMBI	3	4.98(12)	4.08(7)	25%	7.53	4(3)/64(13)	4.19(7%)	0.41	Yasara [50], Snooker, Fleksy [51]		Ligand binding mode prediction with structure-based pharmacophores
UCSF-Shoichet-2	2	3.58(8)	2.6(7)	43%	7.27	3(1)/64(13)	3.89(7%)	0.4	DOCK3.5.54 [54], PLOP [46]	Pose selection	Model selection by VLS enrichment
Stockholm	4	4.36(13)	2.34(7)	41%	6.45	1(1)/64(13)	3.08(5%)	0.4	MacroModel [34], Glide 5.5 [52], AutoDock Vina [64]		
Baylor-Barth	3	5.57(13)	2.65(7)	31%	8.07	3(2)/64(13)	4.61(8%)	0.39	Rosetta [4], Omega [68]	Pose selection	
CDD-CMBI	1	5.28(12)	3.99(7)	15%	8	3(3)/64(13)	4.32(7%)	0.35	Yasara [50], Snooker, Fleksy [51]		Ligand binding mode prediction with structure-based pharmacophores
PharmaDesign	3	4.1(13)	4.12(7)	44%	5.56	2(2)/64(13)	1.71(3%)	0.32	MOE Dock [6]	Model selection, Side-chain and loop modifications	3D pharmacophore search based on experimental anchors
VU-MedChem	3	4.89(13)	3.03(7)	38%	7.84	3(2)/64(13)	3.74(6%)	0.29	GOLD [49], AMBER [44, 69]	Pose selection	

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
Evotec	1	5.95(13)	2.6(7)	35%	7.93	3(1)/64(13)	3.69(6%)	0.27	GOLD [49]	Pose selection	
Baylor-Barth	5	6.23(13)	3.31(7)	31%	6.21	2(1)/64(13)	2.01(3%)	0.26	Rosetta [4], Omega [68]	Pose selection	
PharmaDesign	2	4.44(13)	3.83(7)	44%	6.02	1(1)/64(13)	1.76(3%)	0.25	MOE Dock [6]	Model selection, Side-chain and loop modifications	3D pharmacophore search based on experimental anchors
PharmaDesign	5	4.4(13)	3.96(7)	46%	5.54	1(1)/64(13)	1.18(2%)	0.24	MOE Dock [6]	Model selection, Side-chain and loop modifications	3D pharmacophore search based on experimental anchors
PharmaDesign	1	4.57(13)	4.06(7)	47%	5.34	1(1)/64(13)	0.93(2%)	0.23	MOE Dock [6]	Model selection, Side-chain and loop modifications	3D pharmacophore search based on experimental anchors
Evotec	2	5.83(13)	2.5(7)	47%	7.41	1(1)/64(13)	2.77(5%)	0.2	GOLD [49]	Pose selection	
UMich-Zhang	1	5.74(13)	4.59(7)	2%	9.78	5(1)/64(13)	4.98(8%)	0.19	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
COH-Vaidehi	4	7.56(13)	6.12(7)	34%	7.14	1(1)/64(13)	2.34(4%)	0.17	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
Soochow	1	4.9(13)	1.96(7)	38%	8.09	3(2)/64(13)	3.23(5%)	0.17	Discovery Studio [23]		
UNM	4	4.73(13)	4.72(7)	36%	6.71	1(1)/64(13)	1.78(3%)	0.15	Glide [52]	Pose selection	
Warsaw	3	5.18(13)	3.78(7)	37%	6.72	2(1)/64(13)	1.73(3%)	0.14	Glide 5.6 [52], AutoDock [58]	Pose selection	
UMich-Zhang	3	5.59(13)	4.6(7)	8%	9.78	6(1)/64(13)	4.55(8%)	0.12	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
GaTech	4	8.05(13)	1.9(7)	26%	8.63	3(1)/64(13)	3.43(6%)	0.12	Q-Dock [65], AMBER 8 [44]		
Warsaw	1	5.56(13)	4.43(7)	40%	6.98	2(1)/64(13)	1.57(3%)	0.08	Glide 5.6 [52], AutoDock [58]	Pose selection	
Monash-Sexton-1	1	6.29(13)	3.48(7)	41%	6.7	2(1)/64(13)	1.22(2%)	0.07			
UNM	3	4.75(13)	4.78(7)	30%	7.47	1(1)/64(13)	1.95(3%)	0.06	Glide [52]	Pose selection	
UNM	2	4.56(13)	4.45(7)	18%	6.49	1(1)/64(13)	0.92(2%)	0.05	Glide [52]	Pose selection	
UNM	5	4.26(13)	3.97(7)	39%	6.76	0(0)/64(13)	0.96(2%)	0.02	Glide [52]	Pose selection	
QUB	1	6.83(13)	4.31(7)	27%	8.36	3(1)/64(13)	2.46(4%)	0.01	Glide 5.6 [52], IFD [55], Prime [30], MacroModel [34]	Pose selection	

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
Evotec	3	5.81(13)	2.47(7)	41%	6.98	2(1)/64(13)	1.03(2%)	-0.01	GOLD [49]	Pose selection	
Monash-Sexton-1	5	5.55(13)	3.33(7)	30%	6.34	0(0)/64(13)	0.29(0%)	-0.03			
Warsaw	2	5.56(13)	4.43(7)	40%	7.4	2(2)/64(13)	1.27(2%)	-0.03	Glide 5.6 [52], AutoDock [58]	Pose selection	
Stockholm	1	5.21(13)	2.34(7)	38%	7.33	0(0)/64(13)	1.2(2%)	-0.03	MacroModel [34], Glide 5.5 [52], AutoDock Vina [64]		
Monash-Sexton-1	2	5.3(13)	3.77(7)	39%	6.96	1(1)/64(13)	0.81(1%)	-0.04			
QUB	2	6.68(13)	4.1(7)	31%	9.23	1(1)/64(13)	2.75(5%)	-0.08	Glide 5.6 [52], IFD [55], Prime [30], MacroModel [34]	Pose selection	
Sydney	1	5.19(13)	2.98(7)	31%	8.56	1(1)/64(13)	1.9(3%)	-0.11	Glide 5.5 [52], IFD [55]	Pose selection	
Sydney	5	5.25(13)	2.72(7)	30%	9.94	3(1)/64(13)	3.15(5%)	-0.13	Glide 5.5 [52], IFD [55]	Pose selection	
Sydney	3	5.39(13)	2.88(7)	31%	10.25	3(1)/64(13)	3.29(6%)	-0.15	Glide 5.5 [52], IFD [55]	Pose selection	
QUB	4	6.75(13)	4.34(7)	22%	7.48	0(0)/64(13)	0.44(1%)	-0.18	Glide 5.6 [52], IFD [55], Prime [30], MacroModel [34]	Pose selection	
VU-MedChem	4	6.1(13)	3.54(7)	19%	9.64	2(2)/64(13)	2.37(4%)	-0.2	GOLD [49], AMBER [44, 69]	Pose selection	
Evotec	5	6.18(13)	2.53(7)	33%	7.95	0(0)/64(13)	0.72(1%)	-0.21	GOLD [49]	Pose selection	
Stockholm	5	4.36(13)	2.34(7)	35%	8.87	2(1)/64(13)	1.54(3%)	-0.22	MacroModel [34], Glide 5.5 [52], AutoDock Vina [64]		
Sydney	2	5.75(13)	4.84(7)	20%	7.7	0(0)/64(13)	0.14(0%)	-0.26	Glide 5.5 [52], IFD [55]	Pose selection	
QUB	5	6.64(13)	4.24(7)	34%	8.6	1(1)/64(13)	0.82(1%)	-0.29	Glide 5.6 [52], IFD [55], Prime [30], MacroModel [34]	Pose selection	
Caltech	5	7.57(13)	4.18(7)	27%	10.53	1(1)/64(13)	2.57(4%)	-0.31	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
QUB	3	6.79(13)	4.35(7)	25%	8.59	0(0)/64(13)	0.41(1%)	-0.35	Glide 5.6 [52], IFD [55], Prime [30], MacroModel [34]	Pose selection	
Evotec	4	6.18(13)	2.55(7)	19%	8.56	0(0)/64(13)	0.(0%)	-0.41	GOLD [49]	Pose selection	
Caltech	4	7.34(13)	4.24(7)	36%	8.93	0(0)/64(13)	0.26(0%)	-0.43	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
Sydney	4	5.35(13)	2.83(7)	34%	8.67	0(0)/64(13)	0.(0%)	-0.43	Glide 5.5 [52], IFD [55]	Pose selection	

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
Caltech	3	7.25(13)	3.(7)	26%	8.67	0(0)/64(13)	0.(0%)	-0.43	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
Baylor-Barth	2	6.23(13)	2.53(7)	22%	8.83	0(0)/64(13)	0.06(0%)	-0.44	Rosetta [4], Omega [68]	Pose selection	
Helsinki-Xhaard	3	8.08(13)	5.44(7)	6%	10.32	0(0)/64(13)	0.75(1%)	-0.56	SYBYL-X [24]	Small molecule docking	Manual docking
Helsinki-Xhaard	1	7.86(13)	5.41(7)	5%	10.56	0(0)/64(13)	0.97(2%)	-0.56	SYBYL-X [24]	Small molecule docking	Manual docking
Baylor-Barth	4	4.48(13)	3.3(7)	5%	10.03	0(0)/64(13)	0.11(0%)	-0.62	Rosetta [4], Omega [68]	Pose selection	
UMich-Zhang	5	5.87(13)	4.02(7)	11%	11.13	0(0)/64(13)	1.15(2%)	-0.62	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
UMich-Zhang	2	5.87(13)	4.03(7)	4%	11.29	1(1)/64(13)	1.27(2%)	-0.63	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
Helsinki-Xhaard	2	8.12(13)	5.42(7)	7%	12.82	3(1)/64(13)	2.38(4%)	-0.69	SYBYL-X [24]	Small molecule docking	Manual docking
LenServer	2	19.56(13)	10.56(7)	0%	11.11	0(0)/64(13)	0.(0%)	-0.8			Ligand structure errors
Monash-Sexton-1	3	7.63(13)	2.59(7)	2%	11.16	0(0)/64(13)	0.(0%)	-0.81			
LenServer	1	19.46(13)	10.51(7)	0%	11.36	0(0)/64(13)	0.(0%)	-0.84			Ligand structure errors
UNC	2	7.5(13)	5.4(7)	8%	13.33	2(1)/64(13)	1.8(3%)	-0.86	MedusaDock [38], DMD [61]	Pose selection	
LenServer	4	19.68(13)	10.65(7)	0%	11.62	0(0)/64(13)	0.(0%)	-0.88			Ligand structure errors
Monash-Sexton-1	4	7.21(13)	3.11(7)	3%	11.74	0(0)/64(13)	0.(0%)	-0.9			
UMich-Pogozheva	1	5.94(13)	3.33(7)	1%	12.06	0(0)/64(13)	0.(0%)	-0.95	QUANTA [66], Glide[52]	Ligand docking	
UMich-Lomize	1	6.65(13)	3.48(7)	1%	12.39	0(0)/64(13)	0.(0%)	-1	QUANTA [66], Glide[52]	Ligand docking	
RHUL	4	6.96(13)	5.21(7)	8%	14.36	2(1)/64(13)	1.83(3%)	-1.01	LigDoc [67]		
RHUL	3	6.96(13)	5.21(7)	8%	13.84	2(1)/64(13)	1.25(2%)	-1.02	LigDoc [67]		
UNC	4	7.51(13)	5.68(7)	8%	13.41	0(0)/64(13)	0.32(1%)	-1.1	MedusaDock [38], DMD [61]	Pose selection	
RHUL	1	6.96(13)	5.21(7)	0%	13.09	0(0)/64(13)	0.(0%)	-1.1	LigDoc [67]		

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
RHUL	2	6.96(13)	5.21(7)	0%	13.1	0(0)/64(13)	0.(0%)	-1.11	LigDoc [67]		
UNC	1	7.73(13)	5.57(7)	2%	14.97	2(1)/64(13)	1.6(3%)	-1.14	MedusaDock [38], DMD [61]	Pose selection	
UNC	5	7.76(13)	5.6(7)	7%	15.75	3(1)/64(13)	1.75(3%)	-1.24	MedusaDock [38], DMD [61]	Pose selection	
UNC	3	7.58(13)	5.59(7)	6%	15.82	3(1)/64(13)	1.56(3%)	-1.27	MedusaDock [38], DMD [61]	Pose selection	
UMich-Zhang	4	5.83(13)	4.58(7)	5%	17.32	2(1)/64(13)	1.96(3%)	-1.44	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
RHUL	5	6.96(13)	5.21(7)	2%	16.33	0(0)/64(13)	0.(0%)	-1.6	LigDoc [67]		
Soochow	2	4.9(13)	1.96(7)	6%	17.93	0(0)/64(13)	0.27(0%)	-1.8	Discovery Studio [23]		
LenServer	5	19.48(13)	10.52(7)	0%	18.21	0(0)/64(13)	0.(0%)	-1.89			Ligand structure errors
LenServer	3	19.56(13)	10.54(7)	0%	23.68	0(0)/64(13)	0.(0%)	-2.72			Ligand structure errors
CXCR4/CVX15											
UMich-Zhang	5	6.73(26)	4.11(12)	26%	8.88	7(4)/181(26)	11.64(6%)	2.4	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
COH-Vaidehi	4	9.05(26)	5.19(12)	37%	15.61	11(1)/181(26)	12.96(7%)	1.92	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
COH-Vaidehi	3	9.2(26)	5.37(12)	34%	13.9	6(5)/181(26)	9.51(5%)	1.52	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
UMich-Zhang	4	7.31(26)	5.25(12)	35%	17.06	8(5)/181(26)	10.43(6%)	1.34	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
UMich-Zhang	1	6.4(26)	3.66(12)	41%	10.53	6(1)/181(26)	5.49(3%)	1.21	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
MolLife	2	5.78(25)	3.84(12)	43%	11.85	4(1)/181(26)	6.14(3%)	1.18	GRAMM [70]	Pose selection	Ligand structure errors
MolLife	5	5.7(25)	4.62(12)	39%	14.2	4(4)/181(26)	5.87(3%)	0.89	Enhanced homology modeling [47]	Pose selection	Ligand structure errors
UMich-Zhang	2	6.42(26)	3.68(12)	39%	11.87	3(1)/181(26)	4.36(2%)	0.88	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
COH-Vaidehi	1	9.16(26)	5.34(12)	36%	12.73	2(1)/181(26)	4.76(3%)	0.86	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
QUB	1	6.53(25)	3.41(12)	34%	12.71	3(3)/181(26)	4.71(3%)	0.85	Glide 5.6 [52], IFD [55], Prime [30], MacroModel [34]	Pose selection, Peptide docking	Ligand structure errors
Evotec	4	9.64(26)	5.23(12)	35%	11.64	4(2)/181(26)	3.98(2%)	0.84	GOLD [49]	Pose selection	Ligand structure errors

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
Evotec	3	9.63(26)	5.25(12)	32%	11.39	4(2)/181(26)	3.61(2%)	0.81	GOLD [49]	Pose selection	Ligand structure errors
MolLife	1	5.17(25)	3.86(12)	36%	10.01	0(0)/181(26)	2.44(1%)	0.76	Enhanced homology modeling [47]	Pose selection	Ligand structure errors
UMich-Zhang	3	6.84(26)	3.91(12)	26%	12.71	2(2)/181(26)	3.98(2%)	0.73	TM-align[62], BSP-SLIM [63], SPICKER [13], MacroModel [34]	Pose selection	
Evotec	1	10.1(26)	4.46(12)	28%	11.71	3(3)/181(26)	3.22(2%)	0.71	GOLD [49]	Pose selection	Ligand structure errors
Evotec	2	10.08(26)	5.75(12)	42%	9.96	0(0)/181(26)	1.57(1%)	0.62	GOLD [49]	Pose selection	Ligand structure errors
UNC	1	8.2(26)	3.48(12)	9%	13.82	1(1)/181(26)	2.7(1%)	0.4	MedusaDock [38], DMD [61]	Pose selection	Ligand structure errors
COH-Vaidehi	5	9.3(26)	5.29(12)	32%	14.79	2(1)/181(26)	2.55(1%)	0.27	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
UNC	3	8.25(26)	3.46(12)	11%	13.77	1(1)/181(26)	1.9(1%)	0.27	MedusaDock [38], DMD [61]	Pose selection	Ligand structure errors
RHUL	3	6.99(26)	3.75(12)	10%	14.26	2(1)/181(26)	2.01(1%)	0.24	LigDoc [67]		
Evotec	5	9.69(26)	5.37(12)	28%	11.56	0(0)/181(26)	0.(0%)	0.19	GOLD [49]	Pose selection	Ligand structure errors
UNC	4	7.99(26)	3.5(12)	10%	13.75	0(0)/181(26)	1.1(1%)	0.14	MedusaDock [38], DMD [61]	Pose selection	Ligand structure errors
Stockholm	4	7.09(25)	3.03(12)	13%	14.48	1(1)/181(26)	1.36(1%)	0.11	ZDOCK [53], MacroModel [34]		
Stockholm	3	7.09(25)	3.03(12)	12%	16.35	2(1)/181(26)	2.35(1%)	0.07	ZDOCK [53], MacroModel [34]		
Caltech	1	8.04(26)	4.78(12)	11%	13.32	0(0)/181(26)	0.(0%)	0	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
UNC	5	8.31(26)	3.51(12)	6%	13.33	0(0)/181(26)	0.(0%)	0	MedusaDock [38], DMD [61]	Pose selection	Ligand structure errors
Stockholm	2	6.28(25)	3.04(12)	9%	13.51	0(0)/181(26)	0.(0%)	-0.02	ZDOCK [53], MacroModel [34]		
GaTech	4	10.72(26)	4.27(12)	38%	15.25	0(0)/181(26)	1.01(1%)	-0.03	Q-Dock [65], AMBER 8 [44]	Peptide threading	
UNC	2	8.2(26)	3.53(12)	4%	13.67	0(0)/181(26)	0.(0%)	-0.03	MedusaDock [38], DMD [61]	Pose selection	Ligand structure errors
GaTech	3	10.74(26)	4.26(12)	42%	15.03	0(0)/181(26)	0.85(0%)	-0.04	Q-Dock [65], AMBER 8 [44]	Peptide threading	
GaTech	1	10.73(26)	4.29(12)	36%	15.1	0(0)/181(26)	0.79(0%)	-0.05	Q-Dock [65], AMBER 8 [44]	Peptide threading	

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
MolLife	3	6.17(25)	4.27(12)	43%	14.34	0(0)/181(26)	0.(0%)	-0.11	GRAMM [70]	Pose selection	Ligand structure errors
Helsinki-Xhaard	1	7.22(26)	4.56(12)	11%	14.37	0(0)/181(26)	0.(0%)	-0.11	SYBYL-X [24]	Small molecule docking	Manual docking, Ligand structure errors
Caltech	4	8.99(26)	5.66(12)	5%	19.03	1(1)/181(26)	2.85(2%)	-0.12	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
UMich-Pogozheva	1	6.03(26)	3.41(12)	11%	14.54	0(0)/181(26)	0.(0%)	-0.13	QUANTA [66], Glide[52]	Ligand docking	
UMich-Lomize	1	7.01(26)	3.9(12)	8%	14.56	0(0)/181(26)	0.(0%)	-0.13	QUANTA [66], Glide[52]	Ligand docking	
GaTech	2	10.75(26)	4.17(12)	41%	15.8	0(0)/181(26)	0.67(0%)	-0.15	Q-Dock [65], AMBER 8 [44]	Peptide threading	
QUB	2	6.46(25)	3.13(12)	35%	18.03	1(1)/181(26)	1.72(1%)	-0.21	Glide 5.6 [52], IFD [55], Prime [30], MacroModel [34]	Pose selection, Peptide docking	Ligand structure errors
Caltech	2	7.76(26)	4.12(12)	20%	15.48	0(0)/181(26)	0.(0%)	-0.23	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
PharmaDesign	1	5.01(25)	2.95(12)	49%	16.19	0(0)/181(26)	0.31(0%)	-0.25	MOE Dock [6]	Peptide structure generation, Model selection, Side-chain and loop modifications	3D pharmacophore search based on experimental anchors
COH-Vaidehi	2	9.55(26)	6.19(12)	31%	17.44	1(1)/181(26)	1.04(1%)	-0.26	Glide, Glide XP[52], ZDock [53], GROMACS 4.0.5 [10]	Pose selection	
RHUL	2	6.99(26)	3.75(12)	3%	17.59	0(0)/181(26)	0.96(1%)	-0.29	LigDoc [67]		
Stockholm	5	6.28(25)	3.04(12)	8%	16.83	0(0)/181(26)	0.(0%)	-0.37	ZDOCK [53], MacroModel [34]		
MolLife	4	5.52(25)	3.39(12)	22%	18.03	0(0)/181(26)	0.75(0%)	-0.37	Enhanced homology modeling [47]	Pose selection	Ligand structure errors
Caltech	5	7.48(26)	4.21(12)	7%	18.96	0(0)/181(26)	0.38(0%)	-0.53	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
Caltech	3	7.62(26)	4.94(12)	11%	18.55	0(0)/181(26)	0.(0%)	-0.55	DarwinDock/GenDock (Tanrikulu et al, in preparation), NAMD [3], ZDOCK [53]	Pose selection	MD performed on explicit membrane + water
Soochow	5	8.86(26)	3.93(12)	4%	20.17	0(0)/181(26)	0.96(1%)	-0.56	Discovery Studio [23]		
Soochow	1	8.86(26)	3.93(12)	2%	21.59	0(0)/181(26)	0.(0%)	-0.87	Discovery Studio [23]		
RHUL	1	6.99(26)	3.75(12)	2%	22.2	0(0)/181(26)	0.(0%)	-0.93	LigDoc [67]		

Group	Model	Pocket residue RMSD, A (# residues)	TM pocket residue RMSD, A (# residues)	Fraction predicted pocket	Ligand RMSD (A)	Atomic contacts (#residues) / reference	Correct contact strength (% reference)	Z-score	Software	Manual steps	Comments
Stockholm	1	6.28(25)	3.04(12)	5%	23.19	0(0)/181(26)	0.(0%)	-1.04	ZDOCK [53], MacroModel [34]		
RHUL	4	6.99(26)	3.75(12)	4%	24.38	0(0)/181(26)	0.(0%)	-1.16	LigDoc [67]		
RHUL	5	6.99(26)	3.75(12)	5%	25.26	0(0)/181(26)	0.(0%)	-1.25	LigDoc [67]		
Soochow	4	8.86(26)	3.93(12)	3%	25.48	0(0)/181(26)	0.(0%)	-1.28	Discovery Studio [23]		
Soochow	2	8.86(26)	3.93(12)	3%	26.07	0(0)/181(26)	0.(0%)	-1.34	Discovery Studio [23]		
Soochow	3	8.86(26)	3.93(12)	2%	26.56	0(0)/181(26)	0.(0%)	-1.39	Discovery Studio [23]		

References

1. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-2948.
2. Eswar, N., et al., *Comparative protein structure modeling using Modeller.* Curr Protoc Bioinformatics, 2006. **Chapter 5**: p. Unit 5 6.
3. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.
4. Davis, I.W. and D. Baker, *RosettaLigand docking with full ligand and receptor flexibility.* J Mol Biol, 2009. **385**(2): p. 381-92.
5. Abagyan, R., et al., *Homology modeling with internal coordinate mechanics: deformation zone mapping and improvements of models via conformational search.* Proteins, 1997. **Suppl 1**: p. 29-37.
6. *MOE (The Molecular Operating Environment)*. 2010, Chemical Computing Group Inc.: 1010 Sherbrooke Street West, Suite 910, Montreal, Canada H3A2R7.
7. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structures.* Journal of Applied Crystallography, 1993. **26**(2): p. 283-291.
8. Lu, M., A.D. Dousis, and J. Ma, *OPUS-PSP: An Orientation-dependent Statistical All-atom Potential Derived from Side-chain Packing.* Journal of Molecular Biology, 2008. **376**(1): p. 288-301.
9. Seeliger, D., J. Haas, and B.L. de Groot, *Geometry-Based Sampling of Conformational Transitions in Proteins.* Structure, 2007. **15**(11): p. 1482-1492.
10. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* J Comput Chem, 2005. **26**(16): p. 1701-18.
11. DePristo, M.A., et al., *Ab initio construction of polypeptide fragments: Efficient generation of accurate, representative ensembles.* Proteins: Structure, Function, and Bioinformatics, 2003. **51**(1): p. 41-55.
12. Wu, S. and Y. Zhang, *LOMETS: a local meta-threading-server for protein structure prediction.* Nucleic Acids Res, 2007. **35**(10): p. 3375-82.
13. Zhang, Y. and J. Skolnick, *SPICKER: a clustering approach to identify near-native protein folds.* J Comput Chem, 2004. **25**(6): p. 865-871.
14. Zhang, Y., *I-TASSER server for protein 3D structure prediction.* BMC Bioinformatics, 2008. **9**: p. 40.
15. Zhang, J. and Y. Zhang, *High-resolution protein structure refinement using fragment guided molecular dynamics simulations.* in preparation.
16. Ginalski, K., et al., *3D-Jury: a simple approach to improve protein structure predictions.* Bioinformatics, 2003. **19**(8): p. 1015-1018.
17. Pei, J., B.-H. Kim, and N.V. Grishin, *PROMALS3D: a tool for multiple protein sequence and structure alignments.* Nucleic Acids Res, 2008. **36**(7): p. 2295-2300.
18. Yang, Q. and K.A. Sharp, *Building alternate protein structures using the elastic network model.* Proteins, 2008. **74**(3): p. 682-700.
19. Shen, M.-y. and A. Sali, *Statistical potential for assessment and prediction of protein structures.* Protein Science, 2006. **15**(11): p. 2507-2524.
20. Pogozheva, I.D., et al., *Interactions of Human Melanocortin 4 Receptor with Nonpeptide and Peptide Agonistsâ€ ,â€¡.* Biochemistry, 2005. **44**(34): p. 11329-11341.
21. Chai, B.-X., et al., *Receptorâ^'Antagonist Interactions in the Complexes of Agouti and Agouti-Related Protein with Human Melanocortin 1 and 4 Receptorsâ€ ,â€¡.* Biochemistry, 2005. **44**(9): p. 3418-3431.
22. Guntert, P. and K. Wuthrich, *Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints.* J Biomol NMR, 1991. **1**(4): p. 447-56.
23. *Discovery Studio*. 2010, Accelrys Inc.: San Diego.
24. *SYBYL*. 2008, Tripos-International: St. Louis, MO.
25. Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. **22**(2): p. 195-201.
26. Barth, P., J. Schonbrun, and D. Baker, *Toward high-resolution prediction and design of transmembrane helical protein structures.* Proc Natl Acad Sci U S A, 2007. **104**(40): p. 15682-15687.
27. Wallner, B.r., P. Larsson, and A. Elofsson, *Pcons.net: protein structure prediction meta server.* Nucleic Acids Research, 2007. **35**(suppl 2): p. W369-W374.
28. Bhattacharya, S. and N. Vaidehi, *Computational mapping of the conformational transitions in agonist selective pathways of a G-protein coupled receptor.* J Am Chem Soc, 2010. **132**(14): p. 5205-14.

29. Zhou, H. and Y. Zhou, *Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments.* Proteins, 2005. **58**(2): p. 321-8.
30. *Prime*. 2007, Schrödinger LLC: New York, NY.
31. Soding, J., *Protein homology detection by HMM-HMM comparison.* Bioinformatics, 2005. **21**(7): p. 951-60.
32. Yarov-Yarovoy, V., J. Schonbrun, and D. Baker, *Multipass membrane protein structure prediction using Rosetta.* Proteins, 2006. **62**(4): p. 1010-25.
33. Bray, J.K., R. Abrol, and r. Goddard, W.A, *Method to predict a physiological ensemble of conformations for GPCRs.* To be published, 2010.
34. *MacroModel*. 2005, Schrödinger LLC.: New York.
35. Krieger, E., G. Koraimann, and G. Vriend, *Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field.* Proteins, 2002. **47**(3): p. 393-402.
36. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Research, 2004. **32**(5): p. 1792-1797.
37. Li, Y.Y., T.J. Hou, and W.A. Goddard, 3rd, *Computational modeling of structure-function of g protein-coupled receptors with applications for drug design.* Curr Med Chem, 2010. **17**(12): p. 1167-80.
38. Yin, S., et al., *MedusaScore: an accurate force field-based scoring function for virtual drug screening.* J Chem Inf Model, 2008. **48**(8): p. 1656-62.
39. Ramachandran, S., et al., *Automated minimization of steric clashes in protein structures.* Proteins: Structure, Function, and Bioinformatics, 2010. **79**(1): p. 261-270.
40. Trabanino, R.J., et al., *First principles predictions of the structure and function of g-protein-coupled receptors: validation for bovine rhodopsin.* Biophys J, 2004. **86**(4): p. 1904-21.
41. *Desmond Molecular Dynamics System*. 2010, D. E. Shaw Research: New York.
42. *Maestro-Desmond Interoperability Tools*. 2010, Schrödinger: New York.
43. Bissantz, C., A. Logean, and D. Rognan, *High-throughput modeling of human G-protein coupled receptors: amino acid sequence alignment, three-dimensional model building, and receptor library screening.* J Chem Inf Comput Sci, 2004. **44**(3): p. 1162-76.
44. Case, D.A., et al., *The Amber biomolecular simulation programs.* J Comput Chem, 2005. **26**(16): p. 1668-88.
45. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
46. Zhu, K., et al., *Long loop prediction using the protein local optimization program.* Proteins, 2006. **65**(2): p. 438-452.
47. Nikiforovich, G.V., G.R. Marshall, and T.J. Baranski, *Modeling molecular mechanisms of binding of the anaphylatoxin C5a to the C5a receptor.* Biochemistry, 2008. **47**(10): p. 3117-30.
48. Nikiforovich, G.V., et al., *Modeling the possible conformations of the extracellular loops in G-protein-coupled receptors.* Proteins: Structure, Function, and Bioinformatics, 2010. **78**(2): p. 271-285.
49. Verdonk, M.L., et al., *Improved protein-ligand docking using GOLD.* Proteins, 2003. **52**(4): p. 609-23.
50. Krieger, E., et al., *Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8.* Proteins, 2009. **77 Suppl 9**: p. 114-122.
51. Nabuurs, S.B., M. Wagener, and J. de Vlieg, *A flexible approach to induced fit docking.* J Med Chem, 2007. **50**(26): p. 6507-18.
52. *Glide*. 2010, Schrodinger, LLC: New York.
53. Chen, R., L. Li, and Z. Weng, *ZDOCK: an initial-stage protein-docking algorithm.* Proteins, 2003. **52**(1): p. 80-7.
54. Lang, P.T., et al., *DOCK 6: combining techniques to model RNA-small molecule complexes.* RNA, 2009. **15**(6): p. 1219-30.
55. Sherman, W., et al., *Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects.* J Med Chem, 2005. **49**(2): p. 534-553.
56. Abel, R., et al., *Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding.* Journal of the American Chemical Society, 2008. **130**(9): p. 2817-2831.
57. *Phase*. 2010, Schrödinger, LLC: New York.
58. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.* J Comput Chem, 2009. **30**(16): p. 2785-91.

59. Jain, A.N., *Surflex: A Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine.* J Med Chem, 2003. **46**(4): p. 499-511.
60. Abagyan, R. and M. Totrov, *Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins.* J Mol Biol, 1994. **235**(3): p. 983-1002.
61. Ding, F., et al., *Ab initio folding of proteins with all-atom discrete molecular dynamics.* Structure, 2008. **16**(7): p. 1010-8.
62. Zhang, Y. and J. Skolnick, *TM-align: a protein structure alignment algorithm based on the TM-score.* Nucleic Acids Res, 2005. **33**(7): p. 2302-2309.
63. Lee, H. and Y. Zhang, *BSP-SLIM: A Blind Low-Resolution Ligand-Protein Docking Approach Using Theoretically Predicted Protein Structures.* submitted, 2010.
64. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* J Comput Chem, 2010. **31**(2): p. 455-61.
65. Brylinski, M. and J. Skolnick, *Q-Dock: Low-resolution flexible ligand docking with pocket-specific threading restraints.* J Comput Chem, 2008. **29**(10): p. 1574-88.
66. *QUANTA*. 2005, Accelrys Inc.: San Diego.
67. Vorobjev, Y.N., *Blind docking method combining search of low-resolution binding sites with ligand pose refinement by molecular dynamics-based global optimization.* J Comput Chem, 2010. **31**(5): p. 1080-92.
68. Bostrom, J., J.R. Greenwood, and J. Gottfries, *Assessing the performance of OMEGA with respect to retrieving bioactive conformations.* J Mol Graph Model, 2003. **21**(5): p. 449-62.
69. de Graaf, C. and D. Rognan, *Customizing G Protein-coupled receptor models for structure-based virtual screening.* Curr Pharm Des, 2009. **15**(35): p. 4026-4048.
70. Tovchigrechko, A. and I.A. Vakser, *GRAMM-X public web server for protein-protein docking.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W310-4.

Supplemental Methods: Techniques and Approaches Used for Generation of Complex Models in GPCR Dock 2010

(method descriptions provided by GPCR Dock 2010 participants)

0400

Yasushi Yoshikawa and Toshio Furuya

Research & Development Division
PharmaDesign, Inc.
yoshikawa@pharmadesign.co.jp

1) Homology Modeling

1-1) Alignment and template selection

All Class A GPCR sequences except olfactory receptors were aligned by ClustalW and were refined by hand (about 2,000 sequences). The crystal structure of beta-2 adrenoreceptor (PDB: 2RH1) was employed as a template for modeling of human CXCR4 and human dopamine D3 receptor structure, respectively.

1-2) Initial Homology Modeling

The models of human CXCR4 and human dopamine D3 receptor were built by homology modeling method implemented in MOE 2007 version. The backbone heavy atoms in transmembrane region were fixed. The backbone Phi, Psi dihedral angles of some residues in TM2 helix of CXCR4 model were altered to direct to Trp94 sidechain inside the transmembrane helixes.

1-3) Re-building human CXCR4 and human dopamine D3 receptor models.（10,000 models）

To consider the variety of sidechain rotamers and loop structures, non-conserved residues and E1, E2, E3 loops were generated. Non-conserved residues were replaced to alanine and E1, E2, E3 loops were eliminated. 10,000 models of each receptor were re-built.

1-4) Selecting models for docking

2,000 (CXCR4) or 1,000 (D3 receptor) structures were selected from above models on the basis of potential energy and were used for following docking procedure.

2) Docking

2-1) Ligand preparation

Ligand protonated states were predicted using Chemaxon ACD/ADME suite. Conformations of ligands were generated using MOE "Conformation Import" for CXCR4 small molecule and "Stochastic Search" for D3 ligand, CXCR4 peptide ligand. The CXCR4 peptide ligand was too big to calculate in time, so only the residues from Cys4 to Cys13 was used for the first docking and remained residues were added manually after the docking calculation, and the sidechain structures were generated based on the beta-hairpin structure (PDB: 2AXI).

2-2) Setting experimental anchors

We assigned "Experimental Anchor" of CXCR4 small molecules and dopamine D3 receptor ligand based on site direct mutagenesis (SDM) and structure activity relationship (SAR) (fig. 1). Although Asp97 of human CXCR4 was known as an important residue for ligand binding, due to the sidechain of Asp97 being outside of receptor. we didn't take it into consideration.

2-3) Docking

All of the selected receptor models were used for docking calculation. MOE Dock was used for these dockings. An arbitrary wall restraints were set around GPCR ligand binding sites speculated from known GPCR structures and SDM data. Top 10 high scored poses were output per conformer, which were 1,000,000-2,000,000 poses per target and were used for following method.

3) Narrowing down the poses and optimization

3-1) 3D pharmacophore search

3D pharmacophore search were applied to each output. The 3D pharmacophores were set by referring to each experimental anchors.

3-2) Minimizing the complex structures

The complex models with receptor and selected ligand were generated and minimized. The backbone atoms in transmembrane residues were fixed.

3-3) Selecting models by using the experimental anchor. (CXCR4-small ligand and D3 receptor-eticlopride only)

The distances between ligand and receptor atoms of experimental anchor were measured. The complexes which have appropriate interaction were selected.

3-4) Clustering the docking modes

All output poses of ligands were clustered.

3-5) Selecting suitable cluster by visual inspection

We inspected several of the biggest clusters. The most suitable cluster for experimental data was selected.

3-6) Selecting model from the clusters by visual inspection

We selected the most appropriate docked structure from the selected cluster.

3-7) Optimizing final structure

Some rotamers and loop structures in the selected model were modified manually and energy minimized.

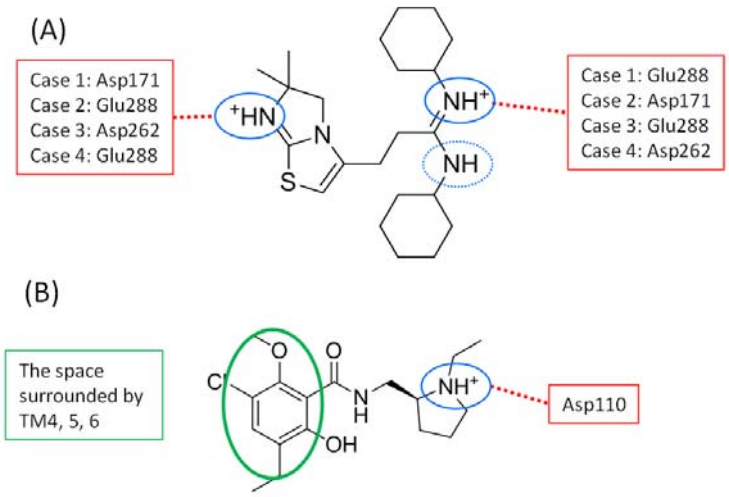


Figure 1. The experimental anchors of (A) CXCR4-small molecule and (B) D3 recptor-eticlpride .

0460

Modeling of Dopamine and CXCR4 ligand-receptor complexes using I-TASSER and BSP-SLIM

Huisun Lee, Ambrish Roy, John Grime, Joseph Rebehmed, Yang Zhang

Center for Computational Medicine and Bioinformatics, University of Michigan,
100 Washtenaw Avenue, Ann Arbor, MI 48109-2218, USA

Receptor structure prediction. The receptor amino acid sequences were first threaded by a meta-threading program, LOMETS [1], through a representative PDB library. Continuous fragments excised from the top-scoring template structures were then used for assembling full-length models by a modified replica-exchange Monte Carlo simulation [2], under the guidance of the spatial restraints collected from the multiple LOMETS templates. The I-TASSER simulation decoys were clustered by SPICKER [3] and the cluster centroids were used as the next round of I-TASSER reassembly. The structures of the lowest energy in the second round assembly simulation were finally selected. The all-atomic models are refined by FG-MD [4], a fragment constrained molecular dynamic simulation procedure, for improving the H-bonding network and removing the steric clashes. The receptor modeling procedure was fully automated.

Ligand-receptor docking. To predict the ligand binding pocket, the unbounded I-TASSER receptor models were searched through a representative ligand-receptor complex library by TM-align [5] to identify structural homologous templates. The binding sites were decided by the clustering of the ligand binding pockets in identified template holo-receptors, combined with the experimental information from similar ligands available in literature [6-9]. The protonated state of small ligands, 1t and eticlopride, was determined by MARVIN (ChemAxon) under acidic pH conditions. In the case of the peptide ligand CVX15, energy-minimization was implemented using MacroModel (Schrödinger, LLC) with distance restraints to allow the formation of a β-sheet structure. Predicted binding pockets in the receptor conformations were docked with low energy ligand conformers using BSP-SLIM [10], resulting in an ensemble of ligand conformations. Spatial clustering of these docked conformations was performed using a modified SPICKER algorithm and ligand poses closest to the cluster centroid were selected from each receptor conformation. In addition, docking poses generated by random translation and rotation in binding site were assessed on the basis of shape and chemical feature complementarity (ionic, ring, hydrophobic and hydrogen-bonding) between ligand and receptor and then candidate complex structures were energy-minimized with an implicit water model using MacroModel with positional restraints to maintain the initial structures of individual TM helices. Final conformations of the ligand-receptor complexes were selected from the high scoring hits based on the balance of the number of ligand clashes with the receptor and the ability to satisfy the experimental restraints.

References

1. Wu, S. and Y. Zhang, *LOMETS: a local meta-threading-server for protein structure prediction.* Nucleic Acids Res, 2007. **35**(10): p. 3375-82.
2. Wu, S., J. Skolnick, and Y. Zhang, *Ab initio modeling of small proteins by iterative TASSER simulations.* BMC Biol, 2007. **5**: p. 17.
3. Zhang, Y. and J. Skolnick, *SPICKER: a clustering approach to identify near-native protein folds.* J Comput Chem, 2004. **25**(6): p. 865-71.
4. Zhang, J. and Y. Zhang, *High-resolution protein structure refinement using fragment guided molecular dynamics.* 2010: p. Submitted.

5. Zhang, Y. and J. Skolnick, *TM-align: a protein structure alignment algorithm based on the TM-score.* Nucleic. Acids Res., 2005. **33**(7): p. 2302-2309.
6. Mansour, A., et al., *Site-directed mutagenesis of the human dopamine D2 receptor.* Eur J Pharmacol, 1992. **227**(2): p. 205-14.
7. Trent, J.O., et al., *Lipid bilayer simulations of CXCR4 with inverse agonists and weak partial agonists.* J Biol Chem, 2003. **278**(47): p. 47136-44.
8. Cox, B.A., et al., *Contributions of conserved serine residues to the interactions of ligands with dopamine D2 receptors.* J Neurochem, 1992. **59**(2): p. 627-35.
9. Coley, C., et al., *Effect of multiple serine/alanine mutations in the transmembrane spanning region V of the D2 dopamine receptor on ligand binding.* J Neurochem, 2000. **74**(1): p. 358-66.
10. Lee, H. and Y. Zhang, *BSP-SLIM: A Blind Low-Resolution Ligand-Protein Docking Approach Using Theoretically Predicted Protein Structures.* 2010: p. Submitted.

1006

Luc Roumen[*], Iwan J.P. de Esch, Rob Leurs, Chris de Graaf[*]
[*]equal contributors

Leiden/Amsterdam Center for Drug Research, Division of Medicinal Chemistry, Faculty of Science, VU University Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands.
C.de.Graaf@vu.nl

Sequence Alignment: The sequences of all CXC Chemokine Receptors were extracted from the Uniprot database.[1] This included the seven CXCR human receptors and the CXCR4 sequence from various species. Sequences were aligned using ClustalW[2] to provide information on sequence conservation and CXCR4 specific motifs. Based on sequence conservation, it was decided that there were likely two cysteine bridges involved receptor structure stabilization, e.g. Cys28-Cys276 (TM1 to ECL3) and Cys109-Cys186 (TM3 to ECL2).
Crystal structure assessment: A structural alignment of several GPCR crystal structures was made using the program MOE.[3] This included the structures of the human A2A Adenosine Receptor (AA2AR, pdbcode 3EML), the human β1-adrenergic receptor (ADRB1, pdbcode 2VT4), the human β2-adrenergic receptor (ADRB2, pdbcode 2RH1), and bovine Opsin structure (pdbcode 3DQB).
TM bundle modelling: Based on sequence similarity of the possible templates to the human CXCR4 sequence (accession number P61073), the ADRB2 structure was chosen as a template for the CXCR4 TM bundle. Manual adaptations were made for the TM bundle on two accounts. Firstly, TM1 was reoriented closer to TM7 to construct a proper packing of the amino acid sidechains between TMs 1, 2 and 7, and to accommodate the hypothesized cysteine bridge between the N-terminus and ECL3. Secondly, molecular dynamic simulations on the TxP fold of TM2 were performed using AMBER[4], comparable to the work by Govaerts et al.[5] and Kellenberger et al.[6] It became clear that a twist of TM2 could be expected, thereby orienting W2.60 and D2.63 to the inside of the TM bundle. This is in line with the observed involvement of D2.63 by Wong et al[7] for other CXCR ligands and earlier work performed on CCR8.[8] A corresponding conformation of TM2 was used for the construction of the final model of the TM bundle.
N-terminus/C-terminus modelling: For modelling of the N-terminus, a structural fit was obtained for the CXCL12 crystal structure co-complexed with a portion of the CXCR4 N-terminus (SDF-1, pdbcode 2K04) on top of the TM bundle. The C-terminus was modelled based on Helix 8 in the ADRB2 structure and its further loop structure was left as a random arrangement.
ECL modelling: The extracellular loops (ECLs) were modelled onto the TM models using the program MODELLER.[9] Two different loops for ECL2 were constructed by the ab initio method of MODELLER and one ECL2 conformation was constructed based on the ECL2 of the Opsin structure (3DQB).
Ligand pose prediction: Based on the basic centres within the ligand 1T, several binding hypotheses were defined with acidic residues in the TM bundle (D2.63, D4.60, D6.58 and E7.39) as also found important by Wong et al.[7] Five different GOLD[10] docking poses for the ligand 1T were selected based on these experimentally guided[7] binding mode hypotheses.[11]
Model Optimization: The final protein-ligand models were optimized using AMBER molecular dynamics simulations. Initial MD simulations were performed using experimentally guided H-bond contraints[7], followed by unconstrained EM/MD to optimize the overall structure.[11]

References
1) The Uniprot Consortium, *Nucleic Acids Res* 2010, 38, D142-148

2) Larkin, MA, Blackshields, G, Brown, NP, Chenna, R, McGettigan, PA, McWilliam, H, Valentin, F, Wallace, IM, Wilm, A, Lopez, R, Thompson, JD, Gibson, TJ, Higgins, DG, *Bioinformatics*, 2007, 23, 21, 2947-2948.
3) MOE (The Molecular Operating Environment) Version 2009.10, Chemical Computing Group Inc. Montreal, Canada, http://www.chemcomp.com.
4) Case, DA, Cheatham III, TE, Darden, T, Gohlke, H, Luo, R, Merz Jr, KM, Onufriev, A, Simmerling, C, Wang, B, Woods, R. The Amber biomolecular simulation programs. *J Comp Chem*, 2005, 26**,** 1668-1688.
5) Govaerts, C, Blanpain, C, Deupi, X, Ballet, S, Ballesteros, JA, Wodak, SJ, Vassart, G, Pardo, L, Permentier, M, *J Biol Chem*, 2001, 276, 16, 13217-13225.
6) Kellenberger, E, Springael, JY, Parmentier, M, Hachet-Haas, M, Galzi, JL, Rognan, D. *J Med Chem*, 2007, 50, 6, 1294-1303.
7) Wong, RS, Bodart, V, Metz, M, Labrecque, J, Bridger, G, Fricker, SP, *Mol Pharmacol*, 2008, 74, 6, 1485-1495.
8) Shamovsky, I, de Graaf, C, Alderin, L, Bengtsson, M, Bladh, H, Borjesson, L, Connolly, S, Dyke, HJ, van den Heuvel, M, Johansson, H, Josefsson, BG, Kristoffersson, A, Linnanen, T, Lisius, A, Mannikko, R, Norden, B, Price, S, Ripa, L, Rognan, D, Rosendahl, A, Skrinjar, M, Urbahns, K, *J Med Chem*, 2009, 52, 23, 7706-7723.
9) Eswar N, Marti-Renom, MA, Webb, B, Madhusudhan, MS, Eramian, D, Shen, M, Pieper, U, Sali, A, *Current Protocols in Bioinformatics*, 2006, 5.6.105.6.30, S15.
10) Verdonk, ML, Cole, JC, Hartshorn, MJ, Murray, CW, Taylor, RD, *Proteins*, 2003, 52, 4, 609-623.
11) De Graaf, C, Rognan, D, *Curr Pharm Des*, 2009, 15, 35, 4026-4048.

1135/4416

Youyong Li*, Tingjun Hou

Institute of Functional Nano & Soft Materials (FUNSOM) and Jiangsu Key Laboratory for Carbon-Based Functional Materials & Devices, Soochow University, Suzhou, Jiangsu 215123, P. R. China.

yyli@suda.edu.cn

1. **Information and data used for modeling**
   - 1.1 **Sequence:** The sequences used for the multiple sequence alignment and modeling were obtained from the pdb files provided by 2010 GPCR dock and the UniProtKB/Swiss-Prot database.
   - 1.2 **Homology templates:** Crystal structure of the β2-adrenergic receptor (β2AR) in complex with carazolol (PDB ID: 2rh1) was used as the homology template.
   - 1.3 **Sequence/template alignments:** sequences were aligned by ClustalW program.
   - 1.4 **Experimental data/restraints:** No experimental data / restraints were used.

2. **Modeling algorithms and tools**
   - 2.1 **Model generation for TM domain and loops:** TM domain was predicted by homology modeling inside Discovery Studio, using the crystal structure of the β2AR in complex with carazolol as a template. Loops were predicted by our CCBTX Monte Carlo method.[1-5] CCBTX method was developed for predicting free energies of polymer chains originally. We combined it with Restraint Generic Protein algorithm for predicting loop structures in proteins, including GPCRs.
   - 2.2 **Model refinement and optimization:** The predicted models were optimized in Discovery Studio package using ChARMm force field.
   - 2.3 **Ligand docking:** Docking of ligands was performed with Discovery Studio package. The ligand has been constructed, optimized and docked at the receptor by "CDOCKER" approach.
   - 2.4 **Using experimentally derived restraints:** No experimentally derived restraint was used.
3. **Criteria for prediction analysis, scoring and ranking**

The docking poses were selected and ranked on the basis of their docking scores and the number of hydrogen bonds formed between the ligand and the receptor.

References

1. Youyong Li, Tingjun Hou, William A Goddard "Computational modeling of structure-function of G protein coupled receptors with applications for drug design." *Current Medicinal Chemistry*, 2010, 17, 1167-1180.
2. Prabal Maiti, Youyong Li, Tahir Cagin, William A Goddard , "Structure of polyamidoamide dendrimers up to limiting generations: A mesoscale description." *J. Chem. Phys.* 2009, 130, 144902.
3. Youyong Li, William A. Goddard III, "Prediction of the structure of G-Protein Coupled Receptor with application for drug design" *Pacific Symposium on Biocomputing,* 2008, 13, 344-353.
4. Youyong Li, Fangqiang Zhu, Nagarajan Vaidehi, William A. Goddard III, et al. "Prediction of the 3D Structure and dynamics of Human DP G-Protein Coupled Receptor bound to an agonist and an antagonist", *J. Am. Chem. Soc.* 2007, 129, 10720 -10731.
5. Youyong Li, William A. Goddard III, "Continuous self avoiding walk with application to the description of polymer chains", *J. Phys. Chem. B*. 2006, 110, 18134-18137.

1178

Michael M. Mysinger*, Dahlia R. Weiss*, John J. Irwin, Brian K. Shoichet

*equal contributors

Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, USA, 94158
shoichet@cgl.ucsf.edu

- **Information and data used for modeling.** Rhodopsin[1], beta-1[2] adrenergic, beta-2[3] adrenergic and adenosine A2A receptor[4] structures were used as templates. PROMALS-3D[5] was used for initial alignment between those template sequences, the competition CXCR4 sequence, all CXCR and CCR sequences, as well as diverse related GPCR sequences identified by Basic Local Alignment Search Tool (BLAST[6]). We refined the alignment using mutational data[7] to bring Asp97 into the binding site, introducing a gap at the top of helix II. Further manual alignment was done particularly the ECLs and the top of helix IV. No experimental restraints were explicitly used in the modeling, however mutational data was used to visually evaluate docked poses.
- **Modeling algorithms and tools.** Modeller v9.8[8] was used for homology model generation. For each of the four crystallographic templates we built 64 homology models. To introduce backbone diversity, elastic network models (ENM) of the known structures were generated using the program 3K-ENM[9]. ENM-based templates were clustered by TM helix *rmsd* (k-means clustering[10]), to give 128 ENM templates per known GPCR structure, for each of these four homology models were generated. In total, we docked to 768 homology models. ECL2 plus the C- and N- terminal residues were removed from the models. Protein Local Optimization Program (PLOP[11]) was used to initially optimize local side chains around the known crystallographic ligands within the CXCR4 modeled binding site. DOCK3.5.54[12] was used to place ligand IT1t in the modeled CXCR4, and a reasonable docked pose was selected manually. Local side chains were then PLOP optimized around this placed 1T1t pose for all 768 models. Known CXCR4 ligands were obtained from ChEMBL04[13]. We automatically generated 35 property-matched DUD[14] style decoys for each ligand (unpublished method). These ligands and decoys were docked, and ligand enrichment was computed[12]. Models were selected for visual inspection based on enrichment, rank of IT1t and a high percentage of ligands interacting Glu288. Five models were selected for a diversity of IT1t poses.

- **References**

1. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M. & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-45.
2. Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C., Henderson, R., Leslie, A. G., Tate, C. G. & Schertler, G. F. (2008). Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486-91.
3. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007). High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-65.
4. Jaakola, V. P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y., Lane, J. R., Ijzerman, A. P. & Stevens, R. C. (2008). The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **322**, 1211-7.

5. Pei, J., Kim, B. H. & Grishin, N. V. (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* **36**, 2295-300.
6. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-10.
7. Wong, R. S., Bodart, V., Metz, M., Labrecque, J., Bridger, G. & Fricker, S. P. (2008). Comparison of the potential multiple binding modes of bicyclam, monocylam, and noncyclam small-molecule CXC chemokine receptor 4 inhibitors. *Mol Pharmacol* **74**, 1485-95.
8. Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815.
9. Yang, Q. & Sharp, K. A. (2009). Building alternate protein structures using the elastic network model. *Proteins* **74**, 682-700.
10. Arthur, D. & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Symposium on Discrete Algorithms (SODA)*.
11. Jacobson, M. P., Friesner, R. A., Xiang, Z. & Honig, B. (2002). On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol* **320**, 597-608.
12. Mysinger, M. M. & Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model* **50**, 1561-73.
13. http://www.ebi.ac.uk/chembl. (2010). ChEMBL 04 Database. EMBL-EBI.
14. Huang, N., Shoichet, B. K. & Irwin, J. J. (2006). Benchmarking sets for molecular docking. *J Med Chem* **49**, 6789-801.

1180

Fiona M. McRobb, Ben Capuano, Ian T. Crosby, David K. Chalmers* and Elizabeth Yuriev*

Medicinal Chemistry and Drug Action, Monash Institute of Pharmaceutical Sciences, Monash University (Parkville Campus), 381 Royal Parade, Parkville, VIC 3052 Australia
David.Chalmers@monash.edu; Elizabeth.Yuriev@monash.edu

Information and data used for modeling

Molecular modeling was performed principally using the Schrödinger software suite; homology models were built with Prime 2.2[1] and docking was carried out with Glide 5.6.[2] The Induced Fit Docking (IFD)[3,4] protocol was used for binding site optimization. Default settings were used, unless stated otherwise.

Homology models of the dopamine D_3 receptor were built using the β_2 (2RH1) crystal structure[5,6] as the template, applying the previously developed multiple sequence alignment,[7] including the point mutation, L119W as crystallized. The T4-lysozyme was removed and ICL3 was not modeled. Highly conserved residues in each TM were anchored and the models were generated in Prime.

Ligand preparation

Previously reported D_3 antagonists were used as the set of active compounds and prepared accordingly.[7] A set of 1000 drug-like decoy compounds with an average molecular weight of 360 g/mol was obtained from Schrödinger (http://www.schrodinger.com).[8] The *S*-isomer of eticlopride was used.

Generation of multiple binding site conformations

Two-hundred models of the D_3 receptor were generated, using IFD to dock a number of selected D_3 antagonists to generate multiple binding site conformations. The docking site was defined as a box 28 Å × 28 Å × 28 Å centered on Asp 3.32, Trp 6.48, Phe 6.52, Tyr 7.43 and up to 20 poses per ligand were collected. In some receptor models, Asp 3.32 and Trp 6.48 were excluded from binding site refinement.

Virtual screening

The decoy set, enriched with active compounds, was docked into each homology model, using Glide XP and ranked by GlideScore. The docking site was defined as a box 28 Å × 28 Å × 28 Å and the centre of the binding site was identified using the coordinates of the centre of carazolol in the β_2 crystal structure (2RH1).

Model analysis

All models were visually inspected and only retained if they met criteria, based on site-directed mutagenesis[9-12] and previous modeling experience.[7]

1. Asp 3.32 and Trp 6.48 in a similar conformation to the β_2 crystal structure.
2. The protonated nitrogen in active compounds aligned to Asp 3.32.
3. Eticlopride and similar compounds in reasonable conformations.

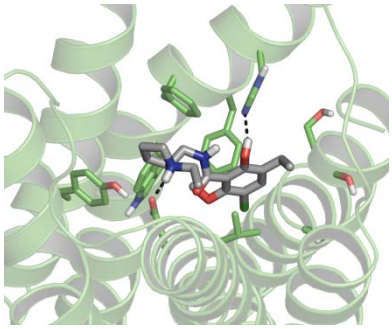


Figure 1: The binding mode of *S*-eticlopride in model 1

Diverse binding modes were selected and ranked by GlideScore, giving the final model ranking (Fig. 1).

References

(1) *Prime*, version 2.2; Schrödinger, LLC: New York, NY, 2010.
(2) *Glide*, version 5.6; Schrödinger, LLC: New York, NY, 2010.
(3) *Schrödinger Suite 2010 Induced Fit Docking protocol;*, Glide, version 5.6; Schrödinger, LLC: New York, NY, 2010: Prime, version 2.2; Schrödinger, LLC: New York, NY, 2010.
(4) Sherman, W.; Day, T.; Jacobson, M. P.*, et al. J. Med. Chem.* **2006,** 49, 534-553.
(5) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.*, et al. Science* **2007,** 318, 1258-1265.
(6) Rosenbaum, D. M.; Cherezov, V.; Hanson, M. A.*, et al. Science* **2007,** 318, 1266-1273.
(7) McRobb, F. M.; Capuano, B.; Crosby, I. T.*, et al. J. Chem. Inf. Model.* **2010,** 50, 626-637.
(8) Friesner, R. A.; Banks, J. L.; Murphy, R. B.*, et al. J. Med. Chem.* **2004,** 47, 1739-1749.
(9) Shi, L.; Javitch, J. A. *Annu. Rev. Pharmacol. Toxicol.* **2002,** 42, 437-467.
(10) Mansour, A.; Meng, F.; Meador-Woodruff, J. H.*, et al. Eur. J. Pharm.-Molec. Ph.* **1992,** 227, 205-214.
(11) Ehrlich, K.; Gotz, A.; Bollinger, S.*, et al. J. Med. Chem.* **2009,** 52, 4923-4935.
(12) Dorfler, M.; Tschammer, N.; Hamperl, K.*, et al. J. Med. Chem.* **2008,** 51, 6829-6838.

1285

Qi Wang, Robert H. Mach, and David E. Reichert*

Division of Radiological Sciences, Mallinckrodt Institute of Radiology, Washington University School of Medicine, 510 South Kingshighway Boulevard, St. Louis, Missouri, 63110

*E-mail: reichertd@wustl.edu. Phone: (314)362-8461. Fax: (314) 362-9940.

Homology Modeling of D3 Receptor. The homology models of D3 receptor were obtained by comparative modeling in the program MODELLER9.2[1], using the crystal structure of the human β2-adrenergic GPCR (pdb code: 2RH1)[2] and the bovine rhodopsin (pdb code: 1F88)[3] as the templates. The sequence of human dopamine receptor was obtained from GeneBank. The sequence alignments were obtained by the automated sequence alignment program ClustalW2[4] with some manual adjustments in the loop regions to maximize homology. The intracellular loop region 3 (IL3) loop was removed from the modeling process. A conserved disulfide bond was built between extracellular loop region 2 (EL2) and TM3 to rigidify the structure. Loop refining was automated by using the "very fast" refinement method within MODELLER. A total of 50 models per receptor subtype were obtained. The model with the lowest DOPE energy calculated by MODELLER was then picked and exported to SYBYL7.2[5] for further manipulation. In SYBYL the bound ligand was modified manually to the D3 antagonist, haloperidol. Hydrogen atoms were added to both the receptor and the ligand, treating the ligand as protonated. The structures were then minimized using the AMBER99 force field and the Gasteiger-Huckel partial charges for the ligand, with the protein backbone atoms fixed in position, to a gradient of 0.05 kcal/(mol Å).

Model Refinement: The resulting model of haloperidol-bound D3 receptor was further refined by multiple molecular dynamics (MD) simulations[6] in explicit membrane/solvent environments. A POPC lipid bilayer membrane patch (75Å 75Å) was generated in VMD1.8.6[7]. The protein was inserted into and combined with the membrane so that the lipid and water molecules overlapping with the protein were deleted. A 5Å layer of TIP3P water molecules was added in the directions above and below the membrane respectively, with an exclusion radii of 3.3Å to solvate the whole system. Counter-ions were also added to produce a salt concentration of 100mM. All the above manipulations were accomplished in VMD1.8.6. The MD simulations were performed using the NAMD2.6 program[8] with the all-atom CHARMM27 force field[9, 10] following a published protocol[11]. Specifically, a total of 10 copies of the system were prepared from the initial starting geometry. While each copy started with the same starting geometry different sampling paths were initiated by different seed numbers; the use of multiple copies allows for better sampling of the system. For each simulation copy, the following five rounds of equilibrations were completed: (i) 2,000 steps of energy minimization for the non-backbone atoms; (ii) five cycles of a 500-step energy minimization with decreasing position restraints (1, 0.8, 0.6, 0.4, 0.2) on the protein Cα atoms; (iii) a gradual increase in the temperature from 50° K to 310° K in 100,000 steps (100ps) of constant volume (NVT ensemble) simulation with restraints (with a force constant of 3 kcal/(mol Å^2)) applied to the protein Cα atoms; (iv) five 40,000 step (40 ps each, 200 ps in total) constant surface area ensemble MD equilibration with decreasing positional restraints on the Cα atoms; (v) 200,000 steps (200 ps) equilibration without any restraints. A total of 500 ps/copy were run for the equilibration steps. Another 1500 ps simulation/copy were run for the production step with the last 500 ps trajectory being collected and used for averaging the structure.

For each simulation copy, the time-averaged structure of the production run was obtained by removing water, lipid and ions from the structure and retaining only the protein and the bound ligand, aligning each frame (1000 frames for a 500 ps trajectory) to the first one and averaging the coordinates. The final averaged structure was obtained by aligning 9 different time-averaged structures from different

copies of simulations to the first one and averaging the coordinates. This structure was then minimized by 500 steps to remove any short range repulsions. The minimized model was saved as the MD-refined, haloperidol-bound homology model of D3 receptor and used for the subsequent studies.
In the simulations, a cutoff of 12 Å was used for non-bonded interactions, and long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) method[12]. Langevin dynamics and a Langevin piston algorithm were used to maintain the temperature at 310K and a pressure of 1 atm. The r-RESPA multiple time step method[13] was employed with a 1 fs time step for bonded atoms, a 2 fs step for short-range non-bonded atoms, and a 4 fs step for long-range electrostatic forces. The bonds between hydrogen and heavy atoms were constrained with the SHAKE algorithm.
Docking of Eticlopride. After removal of haloperidol from the receptor model, the ligand eticlopride was docked into the receptor in GOLD 4.0 (The Cambridge Crystallographic Data Center, Cambridge, UK).[14] Eticlopride was generated in SYBYL and minimized using the MMFF94 force field to a gradient of 0.05 kcal/(mol Å). Since the protonation of the tertiary amine in the ethyl-pyrrolidine ring results in a chiral center, both enantiomers were prepared and used in docking studies. Within GOLD, the ligand was treated as fully flexible, and two residues on the receptor (D3.32 and C 3.36) were also allowed to use side chain conformations from rotamer library. Each ligand was docked 20 times with early termination if the top three poses are within 1.5 Å RMSD. Each pose was ranked according to its GoldScore fitness function. The fitness values from all runs (each enantiomer) were compared and the highest scoring pose was chosen as the binding conformation. Since the top two poses only differ in the flipping of the phenyl ring within the active site, both models were kept as possible conformations.
Further Refinement. Since the sequence of the target in the contest differs a little bit from the sequence we've been using, further refinements of the models were also carried out. Firstly, there is an extra 24 aa length chain at the N-terminus in addition to the model we built; secondly there is a point mutation in TM5, L119W. To deal with the problem, we first used ROSETTA,[15] the ab. inito. prediction method to build models for the 24 aa length chain at the N-terminus since there is no detectable homologous template in pdb to use for homology modeling. 50 models were generated in total for the peptide; the one with the lowest overall score was picked and manually docked to the previously generated D3 model in complex with haloperidol (without any covalent attachment). The purpose is to create a possible 3-dimentional layout for the two. To fuse the peptide to the receptor, both models for the peptide and the receptor were taken as templates in MODELLER and a new target structure was generated with the sequence of the crystal structure, including the point mutation L119W. The new haloperidol bound full sequenced D3 model was again subject to the same kind of MD refinement (3 copies of 2ns simulations).
Eventually the MD refined D3 receptor model was taken as a new template, with the previously eticlopride docked model as another template to generate the refined eticlopride bound full sequenced D3 receptor model in MODELLER.

1. Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A., Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* **2000,** *29*, 291.
2. Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C., High resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **2007,** *318*, 1258.
3. Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M., Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000,** *289*, 739.
4. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G., Clustal W and Clustal X version 2.0. *Bioinformatics* **2007,** *23*, 2947.
5. *SYBYL*, 7.2; Tripos International: St. Louis, MO, USA, 2006.

6. Caves, L. S.; Evanseck, J. D.; Karplus, M., Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci* **1998,** *7*, 649.
7. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** *14*, 33.
8. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J Comput Chem* **2005,** *26*, 1781.
9. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **1998,** *102*, 3586.
10. Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd, Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* **2004,** *25*, 1400.
11. Wang, H. L.; Cheng, X.; Taylor, P.; McCammon, J. A.; Sine, S. M., Control of cation permeation through the nicotinic receptor channel. *PLoS Comput Biol* **2008,** *4*, e41.
12. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* **1993,** *98*, 10089.
13. Tuckerman, M.; Berne, B. J.; Martyna, G. J., Reversible Multiple Time Scale Molecular-Dynamics. *Journal of Chemical Physics* **1992,** *97*, 1990.
14. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R., Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **1997,** *267*, 727.
15. Das, R.; Baker, D., Macromolecular modeling with rosetta. *Annu Rev Biochem* **2008,** *77*, 363.

1576

Gwo-Yu Chuang, Didier Rognan

Structural Chemogenomics
UMR 7200 CNRS-Université de Strasbourg
F-67400 Illkirch, France
chuang@unistra.fr

Modeling The Complex Between The Human Dopamine D3 Receptor And (-) Eticlopride

Information and data used for modeling:
Crystal structures of bovine rhodopsin (PDB ID:1F88)[1] and human β2 adrenergic receptor in complex with carazolol (PDB ID: 2RH1)[2] were used as templates for modeling the human dopamine D3 receptor. Protein sequences were retrieved from the UniProtKB/Swiss-Prot database. The crystal structure of eticlopride (Cambridge Structural Database ID: DOJXEF10)[3] was used as starting conformation of the ligand for docking. Site directed mutagenesis studies[4,5] and structure-activity relationships of orthoethoxybenzamides[6-8] on dopamine D2/D3 receptors were used to guide the initial docking.

Modeling algorithms and tools:
Homology Modeling The initial transmembrane domains (residues 27-56, 63-92, 100-132, 148-170, 185-210, 324-349, 363-383) of the human dopamine D3 receptor was generated with the GPCRmod algorithm[9] starting from a previously modeled human dopamine D2 recepor[10]. Sequence alignments of loops and terminal regions in human dopamine D3, bovine rhodopsin and human β2 adrenergic receptor were separately performed using default settings of CLUSTALW[11]. The N-terminal region (residues 1-26) was modeled based on the crystal structure of bovine rhodopsin, and all the extracellular loops (loop1: residue 93-99; loop2: residue 171-184; loop3: residue 350-362), intracellular loops ((loop1: residue 57-62; loop2: residue 133-147; loop3: residue 211-217, 321-323), and C-terminal region (residue 384-400) were modeled from the crystal structure of the human β2 adrenergic receptor using default settings of MODELLER v9.1[12]. This ligand free structure is subject to energy minimization using SYBYL-X[13].

Docking The side chain of Tyr373 were manually adjusted to open up the antagonist binding cavity. Eticlopride was docked into the binding cavity using GOLD 3.1[14], with the binding site defined as all atoms within 12Å from ASP110 carboxyl oxygen atoms, and using the following constraints: 1) the rotatable bond between the carbonyl carbon and the phenyl moiety of eticlopride is frozen; 2) all protein atoms are fixed, excepted for ASP110 side chain; 3) a distance constraint (minimum separation 2.8Å, maximum separation 3.5Å) between the tertiary amine nitrogen atom of eticlopride and carboxyl oxygen atoms of ASP110. 10 poses ranked by decreasing GoldScore were saved and were inspected manually. The docked pose which satisfied the most experimental constraints was further selected for MD refinement.

MD Refinement The protein-ligand complex was placed in a POPC membrane lipid box solvated with TIP3P water model and subjected to 10 ps of constant volume heating equilibrium and 50 ps of constant pressure constant temperature equilibrium phase using AMBER 8.0[15], with the C_α atoms of the TM residues constrained. A snapshot exhibiting the maximal number of nonbonded contacts between the protein and the ligand was selected from a 1ns MD trajectory and was minimized with AMBER 8.0.

Redocking Eticlopride was docked again into the MD-refined dopamine D3 receptor model using GOLD 3.1 using default settings and two docked poses were selected. Each docked complex was subject to energy minimization using AMBER 8.0.

Criteria for prediction analysis, scoring and ranking: The two docking poses are selected based on the following criteria: 1) H-bond assisted salt bridge between the tertiary amine of eticlopride and ASP110 carboxylic oxygen atoms; 2) maximal number of intermolecular hydrogen bonds; 3) nonbonded contacts with Val86 (conserved in D2, D3 but not D4 subtype) and His349; 5) Other SAR constraints 6) GoldScore. The top ranked model was selected based on the fact that the 3-methyl substituent on the phenyl ring of eticlopride was located in a hydrophobic subpocket lined by Tyr36, Val86, and Trp370.

References:

1. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. Science. 2000 Aug 4;289(5480):739-45.
2. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science. 2007 Nov 23;318(5854):1258-65.
3. de Paulis T, Hall H, Ögren SO, Wägner A, Stensland B, Csöregh I. Synthesis, crystal structure and antidopaminergic properties of eticlopride (FLB 131). European journal of medicinal chemistry 1985 20(3):273-6
4. Lundstrom K, Turpin MP, Large C, Robertson G, Thomas P, Lewell XQ. Mapping of dopamine D3 receptor binding site by pharmacological characterization of mutants expressed in CHO cells with the Semliki Forest virus system. J Recept Signal Transduct Res. 1998 Mar-May;18(2-3):133-50.
5. Mansour A, Meng F, Meador-Woodruff JH, Taylor LP, Civelli O, Akil H. Site-directed mutagenesis of the human dopamine D2 receptor. Eur J Pharmacol. 1992 Oct 1;227(2):205-14.
6. de Paulis T, Kumar Y, Johansson L, Rämsby S, Florvall L, Hall H, Angeby-Möller K, Ogren SO. Potential neuroleptic agents. 3. Chemistry and antidopaminergic properties of substituted 6-methoxysalicylamides. J Med Chem. 1985 Sep;28(9):1263-9.
7. de Paulis T, Kumar Y, Johansson L, Rämsby S, Hall H, Sällemark M, Angeby-Möller K, Ogren SO. Potential neuroleptic agents. 4. Chemistry, behavioral pharmacology, and inhibition of [3H]spiperone binding of 3,5-disubstituted N-[(1-ethyl-2-pyrrolidinyl)methyl]-6-methoxysalicylamides. J Med Chem. 1986 Jan;29(1):61-9.
8. Högberg T, de Paulis T, Johansson L, Kumar Y, Hall H, Ogren SO. Potential antipsychotic agents. 7. Synthesis and antidopaminergic properties of the atypical highly potent (S)-5-bromo-2,3-dimethoxy-N-[(1-ethyl-2-pyrrolidinyl)methyl]benzamide and related compounds. A comparative study. J Med Chem. 1990 Aug;33(8):2305-9.
9. Bissantz C, Logean A, Rognan D. High-throughput modeling of human G-protein coupled receptors: amino acid sequence alignment, three-dimensional model building, and receptor library screening. J Chem Inf Comput Sci. 2004 May-Jun;44(3):1162-76.
10. Bissantz C, Bernard P, Hibert M, Rognan D. Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? Proteins. 2003 Jan 1;50(1):5-25.
11. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.Nucleic Acids Res. 1994 Nov 11;22(22):4673-80.
12. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993 Dec 5;234(3):779-815.

13. SYBYL;Tripos International: St. Louis, MO, 2008.
14. Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J Mol Biol. 1995 Jan 6;245(1):43-53.
15. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. J Comput Chem. 2005 Dec;26(16):1668-88.

1813

Denise Wootten, Patrick Sexton, John Simms

Department of Pharmacology, Monash University, Melbourne, Australia

denise.wootten@med.monash.edu.au

Sequence Alignment

The sequence of the D_3R was aligned against bRho, b2AR and A2aR using ClustalW. The alignment assumed TMs started and ended at the same positions as the b2AR (2RH1).

Homology Modeling

MODELLER9v7 (*1*) was used to generate 200 homology models built using the four available crystal structures, opsin (3CAP), rhodopsin (1GZM), b2AR (2RH1) and A2aR (3ZML) as templates. The resulting ensemble of models was scored using OPUS-PSP (*2*) and the best scoring model was subjected to conformational analysis using CONCOORD. During the CONCOORD runs, 1000 structures were generated and a damping factor of 0.25 was applied to avoid unreasonable side chain geometries. Essential dynamics (*3*) was used to determine the concerted motions of atoms from the ensemble of structures generated from the CONCOORD simulation. A covariance matrix was constructed, using g_covar (GROMACS), which described the correlation of the positional shifts of one atom with those of other atoms and is described as,

$$C_{ij} = ((x_i - x_{io})(x_j - x_{jo}))$$

where x_i and x_j represent the coordinates of atoms i and j in a conformation and x_{io} and x_{jo} represent the average coordinates of the atoms over the ensemble. Rotational and translation motion of the simulation was removed by superimposing each conformation onto an averaged structure, obtained from the ensemble. Subsequent diagonalisation of this matrix yielded a set of eigenvectors that were ordered based on the corresponding eigenvalues. The first 20 eigenvectors were combined and translated back to Cartesian space to generate 500 discrete models of the D3R.

Loop modeling

Extracellular loops (ECL) 1 and 3 were modelled as per the b2AR. RAPPER was used to predict the conformations of ECL2 *ab initio (4)*. The protocol can be separated into 3 discrete steps; (i) *Ab initio* generation of 1000 loop conformations by RAPPER was implemented under the following conditions. Atomic van der Waals radii were scaled down by 25% to ensure good packing and generated loop fragments that contained clashing atoms were discarded. (ii) The sidechain orientations for the predicted backbone conformations were modelled using SCWRL within the environment of the remaining, non-modelled protein. (iii) Generated fragments which exhibited good stereochemistry, as reflected by idealised bond parameters and the absence of clashes with the framework (non-modelled) structure were initial scored using an all-atom statistical potential (scop-e4-allatoms-xray-scores scoring set) as described by (*5*). The ensemble of loop conformations was filtered such that no more than the top 50 models were retained for energy minimisation. Minimisation was performed using the AMBER all-atom forcefield. Minimisation was performed until either convergence or a 0.1 kcal mol-1 cut-off point was reached. Only atoms belonging to either the loop region or the N/C-terminal anchor residues were allowed to move during minimisation. Minimized fragments were subsequently ranked according to the conformational free energy of the loop.

Ligand docking

Eticlopride was docked into each model using suflex and scored. The best scoring model was subjected to a further round of conformational analysis using concord. 1000 models were collected and tested for their enrichment using surflex against a library of amine like ligands seeded with known high affinity known dopamine antagonists. The models which enriched the best were submitted for further evaluation against the crystal structure coordinates.

References

1. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints, J Mol Biol 234, 779-815.
2. Lu, M., Dousis, A. D., and Ma, J. (2008) OPUS-PSP: an orientation-dependent statistical all-atom potential derived from side-chain packing, J Mol Biol 376, 288-301.
3. Amadei, A., Linssen, A. B., and Berendsen, H. J. (1993) Essential dynamics of proteins, Proteins 17, 412-425.
4. DePristo, M. A., de Bakker, P. I., Lovell, S. C., and Blundell, T. L. (2003) Ab initio construction of polypeptide fragments: efficient generation of accurate, representative ensembles, Proteins 51, 41-55.
5. Samudrala, R., and Moult, J. (1998) An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction, J Mol Biol 275, 895-916.
6. Faraldo-Gomez, J. D., Smith, G. R., and Sansom, M. S. (2002) Setting up and optimization of membrane protein simulations, Eur Biophys J 31, 217-227.

2211

Homology Modeling of Human CXCR4 Structure Bound to the Isothiourea Compound and Human Dopamine D3 Structure Bound to Eticlopride.

Dorota Latek, Umesh Ghoshdastider, Slawomir Filipek
University of Warsaw, Faculty of Chemistry, ul. Pasteura 1, 02-093 Warsaw, Poland

Information and data used for modeling:
The CXCR4 and DRD3 protein sequences were aligned with: bovine rhodopsin, human β_2-adrenergic receptor, turkey β_1-adrenergic, human A2A adenosine receptor (protein codes from Protein Data Bank: 1U19, 2RH1, 2VT4, 3EML, respectively) by MUSCLE[1] and CLUSTALW2[2] software. CLUSTALW2 scores were used to choose the most appropriate template (2VT4). Additionally, MUSCLE-derived multiple sequence alignments were prepared from BLAST[3] hits for each target protein sequence. The protonation state of the ligands and placement of disulfide bridges were confirmed by literature search[4,5,6,7].

Algorithm description:
Automatic alignments were manually adjusted to remove gaps inside the TM helices and to preserve disulfide bridges detected experimentally. The homology modeling was performed using the DOPE modeling option in Modeller, followed by slow MD refinement of loops[8]. The best scoring models (according to the DOPE score), three per each target, were subjected to further analysis. The ligands conformations were prepared by Ligprep[9] and Epik[10] protocoles from the Schrodinger Suite. The docking was performed in two ways. First, using the Glide approach[11], and second, using the Autodock 4.2 with Gasteiger charges assignment[12]. The top scoring poses (according to the glide score and the Autodock free energy of binding) were refined in Glide and scored again. The five top scored complexes per each target were submitted. All of them were originally prepared in Glide docking procedure, not in Autodock.

Criteria for prediction analysis, scoring and ranking:
The DOPE (Discrete Optimized Protein Energy) and the DOPE scoring profiles, Glide score, Autodock free energy of binding and visual inspection were used in the assessment.

REFERENCES

[1] Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797.
[2] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and clustal X version 2.0. Bioinformatics 23:2947-2948.
[3] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. J Mol Biol 215:403-410.
[4] Thoma G, Streiff MB, Kovarik J, Glickman F, Wagner T, Beerli C, Zerwes HG (2008) Orally bioavailable isothioureas block function of the chemikine receptor CXCR4 in vitro and in vivo. J Med Chem 51(24):7915-7920
[5] Martelle JL, Nader MA (2008) A review of the discovery, pharmacological characterization, and behavioral effects of the dopamine D2-like receptor antagonist eticlopride. CNS Neurosci Ther 14(3):248-262
[6] Chabot DJ, Zhang PH, Quinnan GV, Broder CC (1999) Mutagenesis of CXCR4 identifies important domains for human immunodeficiency virus type 1 X4 isolate envelope-mediated membrane fusion and virus entry and reveals cryptic coreceptor activity for R5 isolates. J Virol 73(8):6598-6609

[7] Sokoloff P, Giros B, Martres MP, Bouthenet ML, Schwartz JC (1990) Molecular cloning and characterization of a novel dopamine receptor (D3) as a target for neuroleptics. Nature 347(6289):146-151.
[8] Eswar N, Eramian D, Webb B, Shen MY, Sali A (2008) Protein structure modeling with MODELLER. Methods Mol Biol 426:145-159.
[9] LigPrep, version 2.4: Schrödinger, LLC, New York; 2010.
[10] Epik, version 2.1: Schrödinger, LLC, New York; 2010.
[11] Glide, version 5.6: Schrödinger, LLC, New York; 2010.
[12] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J Comput Chem 30:2785-2791

A. CXCR4 Prediction

Andrea Kirkpatrick, Bartosz Trzaskowski, Ravinder Abrol, and William A. Goddard III

A.1 CXCR4: Information and data used for modeling

A.1.1 CXCR4 Data: The human CXCR4 sequence (Swiss Prot # P61073) with the L125W mutation was used for the receptor structure predictions. The transmembrane (TM) helix bundles of human β_2 adrenergic receptor (PDB id: 2rh1) [1], turkey β_1 adrenergic receptor (PDB id: 2vt4, chain B) [2], and human A_{2A} adenosine receptor (PDB id: 3eml) [3] crystal structures served as templates for conformational sampling. Bovine Rhodopsin was excluded due to its low sequence homology to CXCR4. There were two possible alignments for TM regions TM2 and TM4: sequence alignment based on the conserved motifs (D2.50 and W4.50) or the structural (helix shape) alignment based on conserved prolines but with TM2/TM4 placement that positions D2.50/W4.50 at the same z-level with respect to the membrane as the templates. Both alignments were considered for each template to construct the starting TM bundles for receptor structure prediction. No additional experimental data or restraints were used.

A.1.2 Peptide Ligand Data: For the peptide ligand, the provided sequence (Arg-Arg-Nal-Cys-Tyr-Gln-Lys-dPro-Pro-Tyr-Arg-Cit-Cys-Arg-Gly-dPro) was used. This peptide was modeled based on the three protegrin-1 NMR structures obtained from RCSB's PDB: structure 5 from PDB id 1PG1 (with the lowest average RMSD among all 20 structures of 1PG1) [4], structure no 12 from PDB id 1PG1 (with the highest average RMSD among all 20 structures of 1PG1), and monomer A from PDB id 1ZY6 (protegrin-1 dimer bound to membrane)[5]. The residues in these three structures were mapped to the provided peptide sequence (with the exception of Nal and Cit which were replaced by Phe and Arg respectively at this stage and mutated to Nal and Cit during the peptide docking procedure). The first two residues (1 and 2) in protegrin-1 were deleted and the two proline residues (8 and 16) were manually changed to D-prolines using Maestro software (Schrödinger, Inc.). The three peptide conformations were equilibrated by running molecular dynamics (MD) with explicit waters using NAMD [6] for 10ns with the snapshots combined and clustered at 3.7Å diversity resulting in three distinct conformations differing mainly in sidechain orientations. No additional experimental data or restraints were used.

A.2 CXCR4: Modeling algorithms and tools

A.2.1 TM Structure Prediction: The above mentioned starting TM bundles were used to completely sample the helix rotation angle (η) space (using 30° increments: 12^7 ~35 million TM bundle conformations) using the BiHelix/CombiHelix method [7]. This procedure rotated the helices which interact in GPCRs by thirty degrees one pair at a time (12 pairs total). The lowest energy structure from all templates and alignments was used as input for complete local sampling of helix tilt (θ,ϕ) and rotation (η) angles through the SuperBiHelix/SuperCombiHelix method [8] by using: θ (-10°, 0°, 10°), ϕ (-30°, -15°, 0°, 15°, 30°), and η (-30°, -15°, 0°, 15°, 30°). Top six structurally diverse (>2Å relative CαRMSD) protein conformations were selected based on the best average energy-rank criteria described in Sec A.3.1.

A.2.2 Loop Structure Prediction: CXCR4 extracellular loop (ECL) 1, ECL3, intracellular loop (ICL) 1, ICL2, ICL3, and a part of N-terminus (residues 29-36) were generated using the CCBB-Loop method [9-10]. This methodology yielded excellent results when used in recent modeling of the adenosine receptor A_{2A} loops [11]. ECL2 structure was generated using partial homology and similarity to ECL2 of bovine rhodospin [12] and ECL2 of adenosine A_{2A} receptor [3]. The unrestrained part of the N-terminus (residues 1-28) was modeled based on a partial NMR structure of this terminus

complexed with SDF-1α peptide (PDB id: 2K04) [13], and equilibrated using MD with explicit waters that allowed it to sample diverse conformations. The saved MD snapsots were clustered using 4.0Å diversity leading to in 11 distinct conformations for the N-terminus that were attached to the rest of the protein through the Cys28-Cys274 (ECL3) disulfide bond. The C-terminus was modeled based on squid rhodopsin (PDB: 2ZIY) [14].

A.2.3 1t Ligand Docking: Torsionally diverse conformations for 1t ligand were generated by Maestro's MacroModel program[15]. These conformations were clustered at 3.0Å heavy-atom RMSD into 50 diverse families whose family heads were then docked using the DarwinDock/GenDock method [16] to the top 10 CXCR4 TM bundle conformations without loops. The protein-ligand conformations were then selected for MD in the membrane based on their binding energies and diversity.

A.2.4 Peptide Ligand Docking: Docking of the peptide ligand to CXCR4 was performed using the ZDOCK server [17] with the default set of parameters. We docked each of the three peptide conformations obtained above using three different approaches: docking to the N-terminus model, docking to eleven full CXCR4 models (differing in the N-terminus) and docking to one CXCR4 model without the N-terminus present. In all three cases we took the top 500 complexes (by ZDOCK scoring) and performed binding site minimization, neutralization and full complex minimization in the Dreiding force field [18] using the SGB solvation model. Final complexes from each run were selected based on the average of four energy ranks, given in Sec. A.3.2.

A.2.5 Model Relaxation: The predicted complexes for 1t ligand and peptide ligand docked to CXCR4 were embedded in explicit membrane environment and relaxed using MD simulations for 10ns and 5ns respectively. Snapshots saved every 0.5ns during MD were minimized and scored as discussed next.

No experimentally derived restraints were used during structure prediction, ligand docking or relaxation.

A.3 CXCR4: Criteria for prediction analysis, scoring, and ranking

A.3.1 Protein Conformations: The predicted CXCR4 conformations were scored and ranked using the average rank of four different energies associated with each of the TM bundle conformations sampled: charged total energy, charged interhelical energy, neutral total energy, and neutral interhelical energy.

A.3.2 Protein-Ligand Complexes: The predicted CXCR4-1t complexes during docking and after MD relaxation were scored and ranked using vertical binding energy (called *snapbe*), which is simply the energy of the complex minus the energy of the separate ligand minus the energy of the separate protein. The predicted CXCR4-Peptide complexes during docking were scored by using the average rank of four energies (*snapbe*, *interaction*, *total*, and *partial Delphi*) and after MD relaxation by using the average rank of three energies (*snapbe*, *interaction*, and *unified cavity*). *Interaction energy* is the total energy of the ligand interacting with the protein and hence includes ligand strain. *Total energy* is simply the sum of all energies in the complex. *Partial Delphi energy* is the same as *snapbe* except for the ligand desolvation correction. *Unified cavity energy* is the nonbond energy between the ligand and the union of all protein residues within a constant cutoff distance of the ligand in all the complexes being scored.

B. D3 Prediction

Adam Griffith, Soo-Kyung Kim, Ravinder Abrol, and William A. Goddard III

B.1 D3: Information and data used for modeling

The human D3 dopamine receptor sequence (Swiss Prot # P35462) with the L119W mutation was used for the receptor structure predictions. The transmembrane (TM) helix bundles of human β_2 adrenergic receptor (PDB id: 2rh1) [1] and turkey β_1 adrenergic receptor (PDB id: 2vt4, chain B) [2] crystal structures served as templates for conformational sampling. The TM regions for the D3 receptor were identified using the PredicTM method [19]. The predicted TM regions were then aligned to the two templates. There were two possible alignments for TM4: sequence alignment based on the conserved motifs (W4.50) or the structural (helix shape) alignment based on conserved proline (P4.59), but with two possible TM4 placements that position W4.50 or P4.59 at the same z-level with respect to the membrane as the templates. During homology model building, the template helices were extended or truncated as needed to match the D3 TM region predictions. This procedure yielded a total of four D3 templates: two based on β_1 and two based on β_2. No additional experimental data or restraints were used.

B.2 D3: Modeling algorithms and tools

B.2.1 TM Structure Prediction: The SuperBiHelix/SuperCombiHelix method [8] was used to fully sample the TM bundle conformational space spanned by helix tilt angles (θ,ϕ) and rotation angles (η) using the following three overlapping range of angles: a. θ (-10°, 0°, 10°), ϕ (-30°, -15°, 0°, 15°, 30°), and η (-45°,-30°, -15°, 0°, 15°, 30°,45°); b. θ (-20°,-10°, 0°, 10°,20°), ϕ (-45°,-30°, -15°, 0°, 15°, 30°,45°), and η (0°); c. θ (-20°,-10°, 0°, 10°,20°), ϕ (-30°, -15°, 0°, 15°, 30°), and η (-15°, 0°, 15°). Eleven diverse TM bundle conformations (>2Å relative CαRMSD) were selected based on the best average energy-rank, given in Sec.B.3.1.

B.2.2 Loop Structure Prediction: The loops ECL1, ECL2, ECL3, ICL1, and ICL2 were generated using PLOP [20]. The T4L lysozyme piece from the β_2 crystal structure was used in place of ICL3 to connect TM5 and TM6. The C-terminus was generated using our CCBB loop builder program [9-10] which produced only one candidate with helix 8 aligned parallel to the membrane. The N-terminus was modeled using the Bovine Rhodopsin crystal structure. These loops were attached to the six selected protein-ligand complexes by aligning the appropriate helices and minimizing the loops.

B.2.3 Eticlopride Docking: A structure of eticlopride was obtained from the Cambridge Structural Database [21]. The structure was minimized in Jaguar (B3LYP/6-311G**) (Schrödinger, Inc.) and Mulliken charges were obtained for both the protonated and non-protonated forms of the ligand. Additional conformations of the ligand were obtained through the use of MacroModel program [15] in Maestro. These structures were then minimized using Dreiding [18] and clustered at 2.0Å, resulting in 34 diverse ligand conformations which were each then docked using DarwinDock/GenDock [16] to the eleven selected TM bundle conformations into the putative binding site in the TM3/TM5/TM6 region spanning the TM3 conserved Asp, the TM5 Serines, and the TM6 WxPFFxxH motif. The resulting poses for each ligand conformation/protein conformation combination were then pooled together and a diverse set of six docked ligand poses were selected based on multiple energies, given in Sec.B.3.2.

B.2.4 Model Relaxation: The six selected D3-Eticlopride complexes were embedded in explicit membrane environment and relaxed using MD simulations for 4.5ns. Snapshots saved every 0.5ns during MD were minimized and scored as discussed next. Two of the dynamics simulations lost interaction between the ligand and the TM3 Asp and were eliminated from consideration.

No experimentally derived restraints were used during structure prediction or relaxation. During docking, however, one restraint was applied which filtered out any Eticlopride poses that did not form a strong interaction with the conserved Asp on TM3.

B.3 D3: Criteria for prediction analysis, scoring, and ranking

B.3.1 Protein Conformations: The predicted D3 conformations were scored and ranked using the average rank of four different energies associated with each of the TM bundle conformations sampled: charged total energy, charged interhelical energy, neutral total energy, and neutral interhelical energy.

B.3.2 Protein-Ligand Complexes: In DarwinDock the Eticlopride poses were scored and ranked based on the polar component of the interaction energy as the ligand has a protonated amine group. The predicted D3-Eticlopride complexes during docking were scored and ranked using *total*, *snapbe* with ligand strain/solvation, and *unified cavity* energies, and after MD relaxation using *snapbe* with ligand strain/solvation and *unified cavity* energies as defined in Sec. A.3.2. Final ordering of the five deposited D3-Eticlopride complexes was based on a visual analysis of the top structures after MD relaxation.

C. References

1. Cherezov, V., et al., High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science, 2007. 318(5854): p. 1258-65.
2. Warne, T., et al., Structure of a beta1-adrenergic G-protein-coupled receptor. Nature, 2008. 454(7203): p. 486-91.
3. Jaakola, V.P., et al., The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. Science, 2008. 322(5905): p. 1211-7.
4. Fahrner, R.L., et al., Solution structure of protegrin-1, a broad-spectrum antimicrobial peptide from porcine leukocytes. Chem Biol, 1996. 3(7): p. 543-50.
5. Mani, R., et al., Membrane-bound dimer structure of a beta-hairpin antimicrobial peptide from rotational-echo double-resonance solid-state NMR. Biochemistry, 2006. 45(27): p. 8341-9.
6. Phillips, J.C., et al., Scalable molecular dynamics with NAMD. J Comput Chem, 2005. 26(16): p. 1781-802.
7. Abrol, R., J.K. Bray, and W.A. Goddard, 3rd, BiHelix: Towards de novo structure prediction of multispan helical membrane proteins. To be published.
8. Bray, J.K., R. Abrol, and W.A. Goddard, 3rd, SuperBiHelix: Method to predict a physiological ensemble of conformations for GPCRs. To be published.
9. Debe, D.A., et al., Protein fold determination from sparse distance restraints: The Restrained Generic Protein Direct Monte Carlo method. Journal of Physical Chemistry B, 1999. 103(15): p. 3001-3008.
10. Sadanobu, J. and W.A. Goddard, The continuous configurational Boltzmann biased direct Monte Carlo method for free energy properties of polymer chains. Journal of Chemical Physics, 1997. 106(16): p. 6722-6729.
11. Goddard, W.A., et al., *Predicted 3D structures for adenosine receptors bound to ligands: Comparison to the crystal structure.* Journal of Structural Biology, 2010. 170(1): p. 10-20.
12. Palczewski, K., et al., *Crystal structure of rhodopsin: A G protein-coupled receptor.* Science, 2000. 289(5480): p. 739-45.
13. Veldkamp, C.T., et al., *Structural Basis of CXCR4 Sulfotyrosine Recognition by the Chemokine SDF-1/CXCL12.* Science Signaling, 2008. 1(37): p. -.
14. Shimamura, T., et al., *Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region.* Journal of Biological Chemistry, 2008. 283(26): p. 17753-17756.
15. Shelley, J.C., *Comparison of macromodel methodologies for ligand conformation generation.* Abstracts of Papers of the American Chemical Society, 2005. 229: p. U759-U759.
16. Tanrikulu, I.C., et al., *DarwinDock: Ensuring completeness of search for accurate ligand pose prediction in proteins by using an efficient clustering and scoring algorithm.* To be published, 2011.
17. Chen, R., L. Li, and Z.P. Weng, *ZDOCK: An initial-stage protein-docking algorithm.* Proteins-Structure Function and Genetics, 2003. 52(1): p. 80-87.
18. Mayo, S.L., B.D. Olafson, and W.A. Goddard, *Dreiding - a Generic Force-Field for Molecular Simulations.* Journal of Physical Chemistry, 1990. 94(26): p. 8897-8909.

19. Abrol, R., et al., *Prediction of transmembrane helix regions for helical membrane proteins.* To be published, 2011.
20. Jacobson, M.P., et al., *A hierarchical approach to all-atom protein loop prediction.* Proteins-Structure Function and Bioinformatics, 2004. 55(2): p. 351-367.
21. Allen, F.H., *The Cambridge Structural Database: a quarter of a million crystal structures and rising.* Acta Crystallographica Section B-Structural Science, 2002. 58: p. 380-388.
22. Katoh, K., et al., *MAFFT version 5: improvement in accuracy of multiple sequence alignment.* Nucleic Acids Research, 2005. 33(2): p. 511-518.
23. Gasteiger, E., et al., *ExPASy: the proteomics server for in-depth protein knowledge and analysis.* Nucleic Acids Research, 2003. 31(13): p. 3784-3788.
24. Wimley, W.C., T.P. Creamer, and S.H. White, *Experimentally determined hydrophobicity scales for membrane proteins.* Biophysical Journal, 1996. 70(2): p. Tuam1-Tuam1.
25. Bryson, K., et al., *Protein structure prediction servers at university college london.* Nucleic Acids Research, 2005. 33: p. W36-W38.
26. Raghava, G.P.S., *APSSP2 : A combination method for protein secondary structure prediction based on neural network and example based learning.* CASP5, 2002: p. A-132.
27. Pollastri, G. and A. McLysaght, *Porter: a new, accurate server for protein secondary structure prediction.* Bioinformatics, 2005. 21(8): p. 1719-1720.
28. Cole, C., J.D. Barber, and G.J. Barton, *The Jpred 3 secondary structure prediction server.* Nucleic Acids Research, 2008. 36: p. W197-W201.
29. Kam, V.W.T. and W.A. Goddard, *Flat-Bottom Strategy for Improved Accuracy in Protein Side-Chain Placements.* Journal of Chemical Theory and Computation, 2008. 4(12): p. 2160-2169.
30. Bray, J.K. and W.A. Goddard, 3rd, *The structure of human serotonin 2c G-protein-coupled receptor bound to agonists and antagonists.* J Mol Graph Model, 2008. 27(1): p. 66-81.
31. Moustakas, D.T., et al., *Development and validation of a modular, extensible docking program: DOCK 5.* Journal of Computer-Aided Molecular Design, 2006. 20(10-11): p. 601-619.

2560

Prediction of Structural Models of CXCR4 and D3DR receptors and Antagonist Binding sites

Nagarajan Vaidehi, Alfonso Lam, Supriyo Bhattacharya,
Hubert Li, Gouthaman Balaraman, Michiel Niesen

Division of Immunology, Beckman Research Institute of the City of Hope, Duarte, CA.
Email: NVaidehi@coh.org

Methods in Brief:

CXCR4 Structural Models: Of the five top structural models we had submitted for CXCR4 three were homology based models, and two were derived using MembStruk method [1, 2]. The top ranking protein structure is a homology model based on built using Modeler [3] from both adenosine and beta-2-adrenergic receptor crystal structures [4,5] as template. The residues D97 in TM2 and D171 in TM4, which are important in binding of known ligands were not completely inside the binding pocket [6,7]. Hence we optimized the rotational orientation of TM2 and TM4 using the LiTiCon protocol [8]. As part of the LiTiCon protocol, TM2 was rotated in increments of 5° from -30° to 0°. The structure with the minimum in non-bond energy was picked. Similarly, TM4 was optimized by picking the lowest energy structure from a rotation scan in the range of -40° to 0° in increments of 5°. Finally the structure was minimized in CHARMM force field prior to submission. Details of the CXCR4 model based on MembStruk method are given in reference 2. No experimental restraints were used to optimize these models. However, models were selected for ligand docking, by assessment of the position of conserved residues in class A GPCRs and the residues. Intra and extra cellular loops were added using Modeller. The extracellular loop2 and 3 were manually adjusted such that the conserved residues Asp181 and Asp182 form salt bridges with Lys110 and Lys271. respectively. This was our hypothesis based on analogy of our findings on the CXCR3 receptor [Vaidehi and Pease to be published].

D3DR Structural Models: Of the five D3DR structural models, the first three models (by ranking) were derived through homology methods using beta-2-adrenergic receptor (β2-AR) as template [3], while the other two were modeled using Membstruk method [1, 2]. Since the EC-II loop of D3DR shows little homology to β2-AR, this loop was modeled using the *ab-initio* loop modeling procedure in Modeller [3] without using any structural template and an imposed disulfide constraint between Cys103.3 and Cys181 (EC-II). Each of the structures was minimized using the DREIDING force field [9], before subjecting to the docking protocol.

Docking of small molecule ligands: The small molecules 1t for CXCR4, and eticlopride for D3DR were built and optimized using LigPrep module in Maestro. Several ligand conformations were generated outside the protein and docked using Glide XP [Schrodinger Inc]. 10 best docked conformations were retained for each ligand conformation. The docked conformations were clustered using xcluster (Schrodinger Inc) with a cluster distance cutoff of 1.5A. There were two distinct pockets where the docked ligand poses clustered. One was in the TM3, 4, 5 and 6 region as in other biogenic amine receptors. The other pocket was centered around TM1, 2 and 7. Based on the mutation data for cyclam and non-cyclam compounds in CXCR4 (6,7) we selected the binding pocket in the TM1, 2 and 7 region. The top ten best docking models were selected from each cluster using binding energies calculated from all-atom force field as BE = PE (ligand in fixed protein) - PE (ligand in Generalized Born solvation) where BE is the binding energy and PE is the potential energy. The best poses were then selected from the top ten poses using optimal protein-ligand contacts and visual inspection. For eliminating some of the invalid eticlopride docked poses, the mutation data on D2DR [10] (eticlopride

binding is not affected by any of the serine mutations on TM5) was used. The submitted five poses were selected by optimal protein-ligand distances from the top ten docked poses by binding energies.

Docking of the peptide to CXCR4: The initial peptide conformation was obtained through homology modeling from the crystal structure of a related peptide as template [11]. While preparing the peptide structure for docking, all the residues except the charged residues and prolines were mutated to alanine. Multiple conformations of the mutated peptide were generated using torsional sampling of the backbone dihedrals. In the receptor structure, all extracellular loops were removed and the bulky aromatic residues facing the interior of the binding pocket were mutated to alanine. The peptide conformations were then rigidly docked to the CXCR4 structure using Glide (Schrodinger Inc) retaining 10 docked poses from each docking. These docked poses were then clustered by peptide backbone RMSD with a cutoff of 0.5Å. The clusters that were well buried within the transmembrane regions were selected for further analysis and the final poses were chosen based on optimal salt bridges between the ligand and the receptor residues. All mutated residues were then mutated back according to the original sequence and the extracellular loops were added. The loops were then minimized followed by the minimization of the entire complex. We also used Z-dock[12] to dock the peptide, with a distance constraint of 10Å to the center of the binding pocket. The top pose by z-score was taken and the peptide ligand with un-natural amino acids was superimposed to replace the mutated variant. The pose was optimized by a series of MD simulations using GROMACS 4.0.5 [13,14] with the ffG53a6 force field [15]. Energy minimization was performed using a steepest descent algorithm reducing the maximum force on the system to 1000 kJ/mol/nm. The conformation in the binding site was optimized by performing a 100ps annealing simulation on side chain atoms of residues within 5Å of the ligand and the ligand itself. The system was heated from 0K to 300K in the first 25ps and cooled down to 0K in the next 75ps. The same step was repeated with the backbone atoms flexible as well.

References:

1. Trabanino, R. J., Hall, S. E., Vaidehi, N., Floriano, W. B., Kam, V. W. T., and Goddard, W. A..First Principles Predictions of the Structure and Function of G-Protein-Coupled Receptors: Validation for Bovine Rhodopsin. *Biophys. Journal.* **2004***, 86*, 1904-1921.
2. Lam A.R. Bhattacharya S., Patel K., Hall S.E., Mao, A., and Vaidehi, N., The importance of receptor flexibility in binding of cyclam compounds to the chemokine receptor CXCR4, J.Chem Inf. Model. **2010** accepted.
3. Eswar, N., Eramian, D., Webb, B., Shen, M. Y., and Sali, A. Protein structure modeling with MODELLER. Methods Mol. Biol. **2008**, 426, 145–159.
4. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H-J., Kuhn, P., Weis, W. I., Kobilka, B. K., Stevens, R. C., High-resolution crystal structure of an engineered human b2-adrenergic G protein-coupled receptor. *Science*, **2007**, *318*, 1258-1265.
5. Jaakola, V-P, Griffith, M. T., Hanson, M. A., Cherezov V., Chien, E. Y. T., Lane, J. R., Ijzerman, A. P., and Stevens, R. C. .The 2.6 Angstrom Crystal Structure of a Human A_{2A} Adenosine Receptor Bound to an Antagonist. *Science*, **2008**, *322*, 1211-1217.
6. Rosenkilde, M. M., Gerlach, L. O., Hatse, S., Skerlj, R. T., Schols, D., Bridger, G. J. Schwartz, T. W. Molecular Mechanism of Action of Monocyclam versus Bicyclam Non-Peptide Antagonists in the CXCR4 Chemokine Receptor. *J. Biol. Chem*. **2007**, 282 (37), 27354-27365.
7. Wong, R. S. Y., Bodart, V., Metz, M., Labrecque, J., Bridger, G., Fricker, S. P.. Comparison of the Potential Multiple Binding Modes of Bicyclam, Monocylam, and Noncyclam Small-Molecule CXC Chemokine Receptor 4 Inhibitors. *Mol. Pharmacol*. **2008**, *74*, 1485-1495.
8. Bhattacharya S., and Vaidehi N. Computational Mapping of the Conformational Transitions in Agonist Selective Pathways of a G-Protein Coupled Receptor. *J. Amer. Chem. Soc.* **2010**, *132*, 5205-5214.

9. Mayo, S. L., Olafson, B. D. and Goddard, W. A. DREIDING: A generic force field for molecular simulations. *J. Phys. Chem.* **1990**, 94, 8897-8909.
10. Cox, B. A., Henningsen, R. A., Spanoyannis, A., Neve, R. L. and Neve, K. A. Contributions of conserved serine residues to the interactions of ligands with Dopamine D_2 receptors. *J. Neurochem.* **1992**, 59, 627-635.
11. Powers, J.-P. S., Rozek, A. and Hancock, R. E. W. Structure-activity relationships for the β-hairpin cationic antimicrobial peptide polyphemusin I. BBA, 2004, 1698, 239-250.
12. Chen R., and Weng Z. Docking Unbound Proteins Using Shape Complementarity, Desolvation, and Electrostatics. *Proteins.* **2002**, 47(3), 281-294.
13. Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A. E., Berendsen H.J., GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, 26(16), 1701-1718.
14. Hess B., Kutzner C., Van Der Spoel D., Lindahl E., GROMACS 4: Algorithms for Higly Efficient, Load-Balanced, and Scalable Molecular Simulations. *J. Chem. Theory Comput.,* **2008**, 4(3), 435-447.
15. Oostenbrink C., Villa A., Mark A. E., Van Gunsteren W. F., A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.,* **2004**, 25(13), 1656-1676.

2632

Sandeep Pal

Computational chemistry and Molecular Modelling
Evotec (UK) Ltd
Sandeep.pal@evotec.com

Information and data used for modeling

Crystal structures of β2-adrenergic (β2AR) and bovine-rhodopsin (1F88) crystal structure were used for modelling. The mutational data were used from the published references[1-4] .

Model algorithms and tools

The trans-membrane (TM) regions of CXCR4 were determined using the MPEX server[5] . The homology models of the TM regions of CXCR4 and D3 were made using the homology module in MOE using the TM regions of β2-adrenergic (β2AR) and bovine-rhodopsin (1F88) crystal structure as templates. In case of the CXCR4 the homology model of the TM2 region produced an unphysical proline kink. Therefore the TM2 was generated in a two step method. In the first stage an un-kinked TM2 helix was made in MOE. This helix was then kinked using constrained MD simulation using AMBER-94 force field. The TM regions were protonated using the Protonate_3D option in MOE. The modelled TM2 bundles then pass through the next stage of pair-wise helix sampling method.

Pair-wise helix sampling:

The pairwise-helix sampling technique samples the rotational space between two neighbouring helices [Article submitted to J Chem Inf model]. The group of Professor William A. Goddard III has done an exhaustive study in this area[6] . This method of rotational space sampling allows the sampling of 7^7 rotational space for a 7 TM bundle with each helix rotating 7 times around the helical axis. The total bundle interaction energy of the 7 TMs is calculated as follows (**eq 1**):

$$E_{bundle} = \sum_i \sum_j \Delta U_{ij} \qquad (1)$$

where U_{ij} represents the interaction energy between TM_i and TM_j. E_{bundle} values are associated with the rotational configurations of each TM in a bundle. A sorted list of E_{bundle} scores is generated and the bundles with the smallest E_{bundle} scores are analyzed further. The best ranked TM bundles are chosen for docking studies.

Loop modeling. The ECL2 loop was initially modeled using homology to the ECL2 loop of β2AR. The ECL2 cysteine residue and the TM3 cysteine residue were connected to reflect the disulfide bond. Various conformations of the ECL2 were generated using the LowModeMD tool in MOE[7] . During this process the backbone of the TM regions were position restrained. All other ECLs and ICLs were made using homology modeling with β2AR as template.

Docking of Ligands. Prior to the docking the ligands were protonated using the ACD lab software. The protonation state of the nitrogen atom in the imidazothiazole ring was predicted as an acceptor for the IT1t ligand. However this nitrogen atom was reported as a donor in the article by Thoma et al[8] . Both the protonation states were considered while docking IT1t ligand to the CXCR4 model. The docking of the compounds to the TM bundles was performed using GOLD. Flexible docking was performed using the rotamer library of GOLD. In case of the CXCR4 peptide (CVX15) the docked configuration of the peptides was further minimized using low mode molecular dynamic simulations using the MMFF94x force field in MOE.

Criteria for prediction analysis, scoring and ranking:
Top scored (GOLDSCORE and CHEMSCORE) ligand binding modes predicted by GOLD were further analysed using the various site directed mutational analysis data available.
Docking of IT1t in to CXCR4. The top ranked docked poses showed a donor acceptor H-bonding between the nitrogen of the isothiourea group and the Glu 288 in TM7. The nitrogen atom of the imidazothiazole ring either accepted an H-bond from Trp94 (TM2), if considered an acceptor, or donated an H-bond to Asp97 (TM2), if considered a donor. My first two predictions reflect these binding modes.
Docking of CVX15 into CXCR4. The minimized docking poses of the peptides were analysed according to the available mutational data.
Docking of eticlpride to D3. The mutational data show that Asp-110 is important for antagonists binding. The docking poses showed that the basic nitrogen in eticlopride forms a salt bridge with Asp-110. Other important residues important for binding of eticlopride, according to the binding poses, were His349 (TM6), Phe242 (TM6) and Tyr270 (TM7).

CXCR4 sequence:
MEGISIYTSDNYTEEMGSGDYDSMKEPCFREENANFNKIFLPTIYSIIFLTGIVGNGLVILVMGYQKKLRS
MTDKYRLHLSVADLLFVITLPFWAVDAVANWYFGNFLCKAVHVIYTVNLYSSVLILAFISLDRYLAIVHAT
NSQRPRKLLAEKVVYVGVWIPALLLTIPDFIFANVSEADDRYICDRFYPNDLWVVVFQFQHIMVGLILP
GIVILSCYCIIISKLSHSKGHQKRKALKTTVILILAFFACWLPYYIGISIDSFILLEIIKQGCEFENTVHKWISITE
ALAFFHCCLNPILYAFLGAKFKTSAQHALTSVSRGSSLKILSKGKRGGHSSVSTESESSSFHSS

D3 sequence:
MASLSQLSSHLNYTCGAENSTGASQARPHAYYALSYCALILAIVFGNGLVCMAVLKERALQTTTNYLVVS
LAVADLLVATLVMPWVVYLEVTGGVWNFSRICCDVFVTLDVMMCTASIWNLCAISIDRYTAVVMPVHY
QHGTGQSSCRRVALMITAVWVLAFAVSCPLLFGFNTTGDPTVCSISNPDFVIYSSVVSFYLPFGVTVLVYA
RIYVVLKQPLREKKATQMVAIVLGAFIVCWLPFFLTHVLNTHCQTCHVSPELYSATTWLGYVNSALNPVIY
TTFNIEFRKAFLKILSC

Fig 1: The TM regions of the two GPCR sequences are shown in red. The ECL2 regions are shown in green. Other loop regions (N-terminus, ICLs, ECLs and C-terminus) are shown in black.

References

1. Alexander Berchanski; Aviva Lapidot Prediction of HIV-1 entry inhibitors neomycin-arginine conjugates interaction with the CD4-gp120 binding site by molecular modeling and multistep docking procedure. *Biochim. Biophys. Acta* **2007,** *1768*, 2107-2119.

2. Bocker, A.; Sasse, B. C.; Nietert, M.; Stark, H.; Schneider, G. GPCR targeted library design: novel dopamine D3 receptor ligands. *ChemMedChem.* **2007,** *2*, 1000-1005.

3. Sasse, B. C.; Mach, U. R.; Leppaenen, J.; Calmels, T.; Stark, H. Hybrid approach for the design of highly affine and selective dopamine D(3) receptor ligands using privileged scaffolds of biogenic amine GPCR ligands. *Bioorg. Med. Chem.* **2007,** *15*, 7258-7273.

4. Schlotter, K.; Boeckler, F.; Hubner, H.; Gmeiner, P. Fancy bioisosteres: novel paracyclophane derivatives as super-affinity dopamine D3 receptor antagonists. *J. Med. Chem.* **2006,** *49*, 3628-3635.

5. Jayasinghe, S.; Hristova, K.; White, S. H. Energetics, stability, and prediction of transmembrane helices. *J. Mol. Biol.* **2001,** *312*, 927-934.

6. Goddard, W. A., III; Abrol, R. 3-Dimensional structures of G protein-coupled receptors and binding sites of agonists and antagonists. *J. Nutr.* **2007,** *137*, 1528S-1538S.

7. Paul Labute LoWModeMD-Implicit Low-Mode Velocity Filtering Applied to Comformational Search of Macrocycles and Protein Loops. *J Chem Inf Model* **2010,** *50*, 792-800.

8. Gerhard Thoma et al Orally Bioavailable Isothioureas Block Function of the Chemokine Receptor CXCR4 In Vitro and In Vivo. *J Med Chem* **2008,** *51*, 7915-7920.

Blind docking method: analysis of protein accessible surface and refinement of binding models by molecular dynamics

Yrii Vorobjev[2,] Natalia Bakulina[2] and Victor Solovyev [1]

[1]Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX,UK; [2]Softberry Inc., 116 Radio Circle, Suite 400, Mount Kisco, NY 10549, USA
victor@cs.rhul.ac.uk

Template selection for homology modeling of GPCR proteins was based on sequence identity and literature data. The candidate template structures of GPCR proteins were found by ***Ffold*** program. The structure of turkey beta1 adrenergic receptor with stabilizing mutations and bound cyanopindolol (PDB ID 2VT4) was selected as a template for human dopamine D3. The crystal structure of bovine rhodopsin at 2.2 angstroms (PDB ID 1U19) was used as a template for homology modeling of CXCR4. Template-target models of each GPCR were build by Modeller.

A hierarchical blind docking method ***LigDoc*** consists of two steps. The first step is an analysis of protein accessible surface to identify pockets and clefts. The surface analysis is performed for a given rigid protein structure. A probe sphere of an average atomic size rolls over the protein surface to get the probe contacts with protein atoms. The set of protein atom - probe contacts is transformed into a manifold of extended dots, which are clustered and the centers of clusters represent a set of coordinates of low-resolution binding sites. The number of extended dots in the cluster gives the contact score of the low-resolution binding site. It was found that the native binding site centers are among of the 6 – 10 top ranking low-resolution binding sites.

The second stage of hierarchical blind docking method is a refinement of the top ranking low-resolution binding sites by a global optimization of position, orientation of the ligand and the protein binding site and conformations via the method of molecular dynamics coupled with temperature annealing and ligand-protein force field reversible deformation. Optimal protocol of the global optimization was found for training set protein-ligand complexes. The ligand-binding mode with the lowest potential energy of protein ligand interaction is selected as the best binding mode for the hierarchical blind docking method. Five binding modes have been selected for each complex on the base of analysis of energy of ligand protein interactions.

3041

Thijs Beuming[1], Stefano Costanzi[2], Lei Shi[3], Chris Higgs[1], Noeris Salam[1], Dmitry[1] Lupyan[1], Woody Sherman[1]

1. Schrödinger Inc., New York, New York 10036
2. Laboratory of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, DHHS, Bethesda, MD 20892, USA
3. Department of Physiology and Biophysics and HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, Cornell University, 1300 York Avenue, New York, NY 10021, USA.

An initial sequence alignment of the human D_3 receptor with available GPCR templates from the Protein Data Bank was obtained using the GPCR alignment module in Prime (version 2.2, Schrödinger, LLC, New York, NY, 2010). This improves the alignment of the conserved motifs (fingerprints) that characterize each of the transmembrane (TM) helices. In select cases, the TM alignment was manually adjusted. Gaps in the extra- and intra-cellular loops were consolidated into a single gap per loop, and positioned where insertions or deletions seemed compatible with the structure of the template. The structure of the β_2-adrenergic receptor (β_2-AR, PDB code 2RH1) was used as an initial template for the generation of the D_3-model, using the Prime *bldstruct* module with default values. The alignment between target and template is shown in Figure 1. Into this starting model several modifications were introduced based on structural information from other templates and available experimental data. These modifications included:

1) The adjustment, when needed, of rotameric states of residues of the D_3 receptor that were not conserved in the β_2-AR, but were conserved in another GPCR with an experimentally elucidated structure, i.e. bovine and squid rhodopsin, the human adenosine A_{2A} receptor, or the turkey β_1-adrenergic receptor (β_1-AR).

2) The second intra-cellular loop (IL2) was modeled as an ideal α-helix (residues 137 to 144) in an orientation identical to the structure of β_1-AR. To improve the connection between IL2 and TM4 all residues C-terminal of the conserved tyrosine (139 to 149) were minimized.

3) In addition to the conserved disulfide-bond between EL2 and TM3, a second disulphide bond (present in the A_{2A} receptor, and expected in D_3 based on sequence information) was introduced in the third extra-cellular loop through a constrained minimization.

Prime loop refinements were then carried out to identify low-energy conformations of extra-cellular loops 1, 2, and 3. Based on the high structural conservation of IL1 in all available structures, this loop was not optimized. IL2 was modeled as described above IL3 was not modeled due to its length and low structural conservation. For EL1, conserved Trp96 was kept in a position similar to the □$_2$-AR template and flanking segments 92-95 and 97-100 were optimized separately. For EL2, only the portion of the loop downstream of the conserved disulphide bridge was built. In the refinement, only the solutions in which Ile183 adopted an inward facing conformation of were accepted, based on substituted-cysteine accessibility data of the highly homologous dopamine D2 receptor (Shi and Javitch 2004).

This optimized model was used for subsequent docking studies, resulting in solutions 1 through 4 (see below).

Eticlopride was docked into the initial model using Induced-Fit Docking (IFD, (Sherman et al. 2006)). This produced two low-energy solutions with acceptable orientations, namely, forming a salt-bridge with the conserved Asp110 (3.32), and having a significant overlap with the carazolol binding mode in β_2-AR, as quantified by a structural interaction fingerprint score (SIFt, (Deng et al. 2004)) . These conformations differed in the orientation of the aromatic ring and the basic group. Two additional hybrid conformations were manually generated based on a simple symmetry operation around the amide bond, and these four poses were submitted for further analysis.

The four poses were evaluated using WaterMap (Abel et al. 2008) in order to get a better estimate for the most stable pose based on the energy of displacing waters (i.e. hydration sites). WaterMap was run on the D3 model and poses were scored by summing over the free energy of hydration sites that were displaced. Two very high energy hydration sites were detected at the TM5 and TM6 interface. The pose ultimately reported in model 1 had the most favorable WaterMap score due to an effective displacement of these two high energy waters by the ethyl- and chloro-substituents on the aromatic ring (see Figure 2).

Finally, a pharmacophore hypothesis of 30 D3 antagonists and inverse agonist obtained from various sources (Holmes et al.; Dubuffet et al. 1999; Varady et al. 2003; Newman et al. 2005; Geneste et al. 2006; Micheli et al. 2008) was prepared to validate the predicted binding mode. Three high-scoring 4-point hypotheses were calculated using Phase (version 3.2, Schrödinger, LLC, New York, NY, 2010). By definition, each hypothesis contained a specific alignment of an eticlopride-conformation to the hypothesis. The three hypotheses were then matched with the four IFD-derived poses by superimposing the eticlopride pose aligned to the pharmacophore with the docked poses. In the case of model 1, but not models 2, 3 and 4, there was a good overlap between the pharmacophore features from one of the hypotheses and the corresponding elements in the docked eticlopride, and no overlap of any of the features with the receptor. Encouragingly, all ligands aligned to this hypothesis had a reasonable fit into the D3 binding-site, when superimposed along with the eticlopride conformation.

To investigate additional modeling strategies, one alternative model (model 5) was prepared with variations of the loop conformations, including:
1) A model of EL2 where in addition to Ile183, Asn185 was also brought into an inward facing conformation by running a restrained molecular dynamics simulation using Desmond (Desmond Molecular Dynamics System, version 2.4, D. E. Shaw Research, New York, NY, 2010).
2) A model of IL2 based directly on the β_2-AR template, which is in an unfolded conformation.
3) A model of EL3 based on rhodopsin in which a disulfide bridge was built between residues equivalent to rhodopsin His278 and Ser281.

The resulting IFD docking into this model produced a pose that was relatively similar to model 1.

Abel, R., Young, T., Farid, R., Berne, B.J., and Friesner, R.A. 2008. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J Am Chem Soc* **130:** 2817-2831.

Deng, Z., Chuaqui, C., and Singh, J. 2004. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. *J Med Chem* **47:** 337-344.

Dubuffet, T., Newman-Tancredi, A., Cussac, D., Audinot, V., Loutz, A., Millan, M.J., and Lavielle, G. 1999. Novel benzopyrano[3,4-c]pyrrole derivatives as potent and selective dopamine D3 receptor antagonist. *Bioorg Med Chem Lett* **9:** 2059-2064.

Geneste, H., Backfisch, G., Braje, W., Delzer, J., Haupt, A., Hutchins, C.W., King, L.L., Kling, A., Teschendorf, H.J., Unger, L., et al. 2006. Synthesis and SAR of highly potent and selective

dopamine D(3)-receptor antagonists: 1H-pyrimidin-2-one derivatives. *Bioorg Med Chem Lett* **16:** 490-494.

Holmes, I.P., Blunt, R.J., Lorthioir, O.E., Blowers, S.M., Gribble, A., Payne, A.H., Stansfield, I.G., Wood, M., Woollard, P.M., Reavill, C., et al. 2010. The identification of a selective dopamine D2 partial agonist, D3 antagonist displaying high levels of brain exposure. *Bioorg Med Chem Lett* **20:** 2013-2016.

Micheli, F., Bonanomi, G., Braggio, S., Capelli, A.M., Celestini, P., Damiani, F., Di Fabio, R., Donati, D., Gagliardi, S., Gentile, G., et al. 2008. New fused benzazepine as selective D3 receptor antagonists. Synthesis and biological evaluation. Part one: [h]-fused tricyclic systems. *Bioorg Med Chem Lett* **18:** 901-907.

Newman, A.H., Grundt, P., and Nader, M.A. 2005. Dopamine D3 receptor partial agonists and antagonists as potential drug abuse therapeutic agents. *J Med Chem* **48:** 3663-3679.

Sherman, W., Day, T., Jacobson, M.P., Friesner, R.A., and Farid, R. 2006. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **49:** 534-553.

Shi, L., and Javitch, J.A. 2004. The second extracellular loop of the dopamine D2 receptor lines the binding-site crevice. *Proc Natl Acad Sci U S A* **101:** 440-445.

Varady, J., Wu, X., Fang, X., Min, J., Hu, Z., Levant, B., and Wang, S. 2003. Molecular modeling of the three-dimensional structure of dopamine 3 (D3) subtype receptor: discovery of novel and potent D3 ligands through a hybrid pharmacophore- and structure-based database searching approach. *J Med Chem* **46:** 4377-4392.

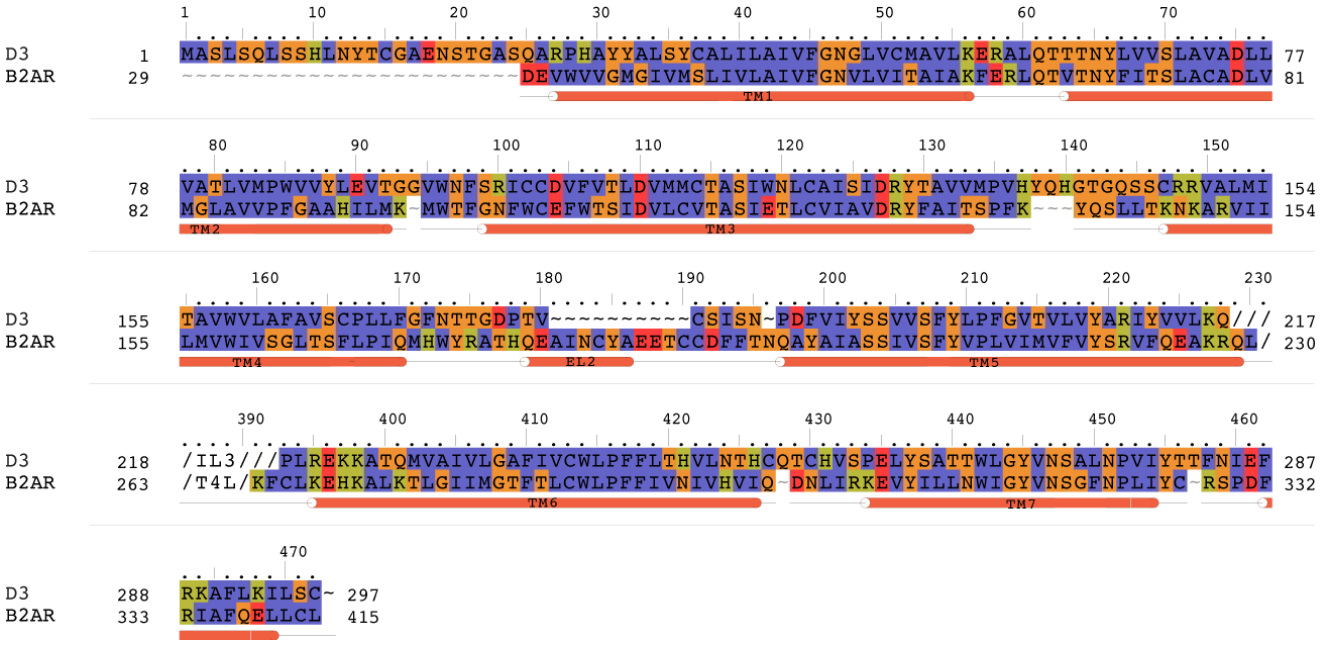


Figure 1. Sequence alignment of the human D3 receptor with the β_2-adrenergic receptor. Helices in the template are indicated with orange bars. The T4 lysosyme (B2AR) and IL3 (D3) sequences have been omitted from the alignment. Figure prepared using the Multiple Sequence Viewer in Maestro.

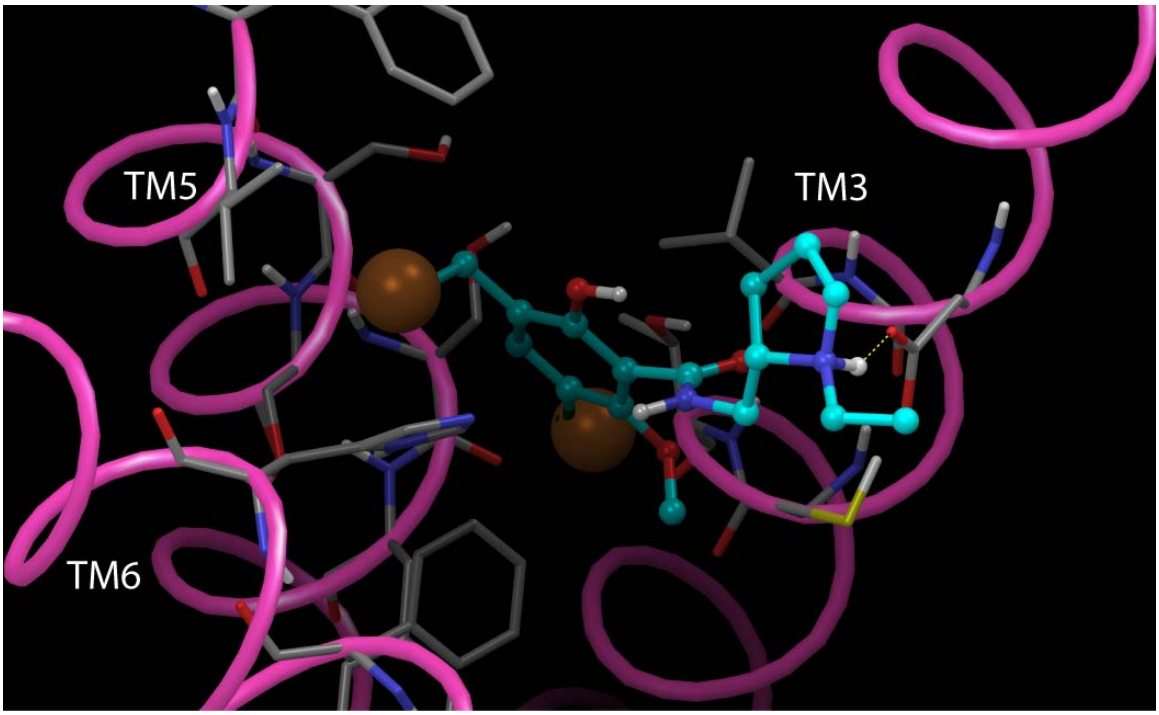


Figure 2. Effective displacement of two predicted high-energy hydration sites by eticlopride. Residues within 4Å of the hydration sites are shown. TMs 1,2,4 and 7 are omitted for clarity.

3532

Feng Ding, Pradeep Kota, Srinivas Ramachandran, and Nikolay V. Dokholyan

University of North Carolina
dokh@med.unc.edu

Structural modeling of GPCRs

To obtain template structures for homology model building, we performed three iterations of PSI-BLAST against the PDB database with the primary sequence of D3 (and CXCR4) as input. We chose the structure of human β1-adrenergic receptor (PDB ID: 2VT4) as a template among the high scoring structures obtained using PSI-BLAST (PDB ID: 2Z73, 3D4S for D3 and 1GZM, 2R4R for CXCR4). We aligned using CLUSTALW (www.ebi.ac.uk/clustalw), the primary sequence of D3 (and CXCR4) to that of the human β1-adrenergic receptor to identify stretches of amino acids that are not covered by the identified structural template. We used Medusa to build a homology model of D3 (and CXCR4) by replacing the side chain of every amino acid in the template (2VT4) to its corresponding amino acid from the alignment (1). We performed fixed backbone redesign to minimize the deviation of the homology model from the structural template. We then relaxed the model using Chiron to resolve steric clashes due to side chain substitutions (2). To account for the residues that do not feature homology with the template, we used the standalone version of psipred (3) and predicted the secondary structure of the fragments that are absent from the template. We then used all-atom discrete molecular dynamics (DMD) to fold the fragments *ab initio* where the transmembrane helices were constrained while the loops were allowed to sample different configurations (4).

Small molecule ligand pose prediction

From the homology model of both D3 and CXCR4, we identified the conserved ligand-binding pocket. We performed the flexible docking simulations using MedusaDock (5), where the conformational flexibility of both ligand and receptor sidechains are simultaneously modeled. For each case, we generated 2000 poses and rank-ordered these poses according to both the MedusaScore (6) and the total energy of the ligand-receptor complex (5). We selected the top 100 poses for clustering analysis. The representative structures of the largest five clusters were selected as the predicted poses.

Cyclic peptide docking

To predict the peptide docking pose, we performed DMD simulations (4). During the simulation, the TM helixes of CXCR4 were constrained while the extracellular loops and peptide were allowed to move freely. We performed replica-exchange DMD simulations to enhance the sampling. Similar to small molecular docking, we ranked the poses according to both binding energy and total energy and clustered the top ranked poses (200) based pair-wise RMSD. The representative structures of the largest five clusters were selected as the predicted poses

References:

1. Ding F., and Dokholyan, N.V. Emergence of protein fold families through rational design. *PLoS Comput Biol.* 2:e85 (2006).
2. Ramachandran, S., Kota, P., Ding, F. and Dokholyan, N.V. Automated minimization of steric clashes in protein structures", *Proteins: Structure, Function and Bioinformatics*, *in press* (2010)
3. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292: 195-202 (1999).
4. Ding, F., Tsao, D., Nie, H. F. and Dokholyan, N. V. Ab initio folding of proteins with all-atom discrete molecular dynamics. *Structure* 16, 1010-1018 (2008).

5. Ding, F., Yin, S., and Dokholyan, N. V. "Rapid flexible docking using a stochastic rotamer library of ligands", *Journal of Chemical Information and Modeling*, 50:1623-1632 (2010)
6. Yin, S., Biedermannova, L., Vondrasek, J., and Dokholyan, N. V., "MedusaScore: An accurate force-field based scoring function for virtual drug screening", *Journal of Chemical Information and Modeling*, 48:1656-1662 (2008)

3646

Jens Carlsson*, Ryan G. Coleman*, Hao Fan, Avner Schlessinger, John J. Irwin, Andrej Sali, Brian K. Shoichet

Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, USA, 94158 *equal contributors

Information and data used for modeling

The β_1[1] and β_2[2] adrenergic receptor structures were used as templates in the modeling. The two template sequences, the D_3 sequence, and all dopamine receptor sequences in UniProt[3] were used to construct the initial alignment. Two disulfide bridges (residues 103-181 and 355-358) were identified based on structural homology to the available crystallographic structures. Mutational data was not used explicitly in model building or docking.

Modeling algorithms and tools

PROMALS3D was used to construct the initial alignment[4]. The resulting alignment was edited only for ECL2. To achieve conformational diversity of the backbone in the template structures, a library of backbone templates were built using the elastic network model program 3K-ENM[5]. This resulted in an additional 350 and 360 templates for the β_1 and β_2 adrenergic receptor structures, respectively.

Modeller version 9.8[6] was used to generate all homology models. The first 24 N-terminal residues were not modeled. Based on the results of initial models, a number of binding site residues were subjected to additional chi-angle restraints in the generation of structures. For each of the templates from x-ray crystallography and 3K-ENM, 250 homology models were generated, then scored using DOPE[7]. 64 models for each x-ray structure and 4 models from each normal mode template were selected for further evaluation. In total, this resulted in 2968 models, which were further assessed using molecular docking.

All molecular docking calculations were carried out using DOCK3.5.54[8, 9]. Each homology model was evaluated based on their ability to enrich known ligands among a large number of decoys. Dopamine D_3 ligands with an affinity better than 10μM were extracted from ChEMBL[10]. Three different decoy databases were used: (1) Property-matched decoys [11] (2) A random set of lead like compounds from the ZINC database (3) A random set of non-GPCR ligands from ChEMBL.

Criteria for prediction analysis, scoring and ranking

Homology models were ranked based on their ability to enrich known dopamine receptor ligands. In addition, at least 60% of the best scoring ligands had to form a charge-charge interaction with Asp110 (Asp3.32). The 100 best models were examined manually by inspecting the binding poses of docked ligands. The conformational sampling of eticlopride was constrained to conserve the internal hydrogen bonds observed in the Cambridge Structural Database[12]. From these, 5 models with varying binding poses for eticlopride were chosen.

References

1. Warne, T., et al., *Structure of a beta1-adrenergic G-protein-coupled receptor.* Nature, 2008. **454**(7203): p. 486-491.

2. Cherezov, V., et al., *High-Resolution Crystal Structure of an Engineered Human β2-Adrenergic G Protein–Coupled Receptor.* Science, 2007. **318**(5854): p. 1258-1265.
3. The UniProt Consortium, *The Universal Protein Resource (UniProt) in 2010.* Nuc Acid Res, 2010. **38**: p. D142-D148.
4. Pei, J., B.H. Kim, and N.V. Grishin, *PROMALS3D: a tool for multiple protein sequence and structure alignments.* Nuc Acid Res, 2008. **36**(7): p. 2295-2300.
5. Yang, Q. and K.A. Sharp, *Building alternate protein structures using the elastic network model.* Proteins: Struct. Funct. Bioinf., 2009. **74**(3): p. 682-700.
6. Sali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints.* J Mol Biol, 1993. **234**: p. 779-815.
7. Shen, M.-Y. and A. Sali, *Statistical potential for assessment and prediction of protein structures.* Prot Sci, 2006. **15**: p. 2507-2524.
8. Lorber, D.M. and B.K. Shoichet, *Hierarchical Docking of Databases of Multiple Ligand Conformations.* Curr Topics Med Chem, 2005. **5**(8): p. 739-749.
9. Kuntz, I.D., et al., *A geometric approach to macromolecule-ligand interactions.* J Mol Biol, 1982. **161**(2): p. 269-288.
10. *ChEMBL*. Available from: http://www.ebi.ac.uk/chembl.
11. Huang, N., B.K. Shoichet, and J.J. Irwin, *Benchmarking Sets for Molecular Docking.* J Med Chem, 2006. **49**(23): p. 6789-6801.
12. Allen, F.H., *The Cambridge Structural Database: a quarter of a million crystal structures and rising* Acta Cryst B, 2002. **58**: p. 380-388.

3682

Irina Tikhonova

School of Pharmacy
Medical Biology Centre
Queen's University Belfast
United Kingdom
i.tikhonova@qub.ac.uk

Homology Modelling of the D3 and CXCR4 Receptors

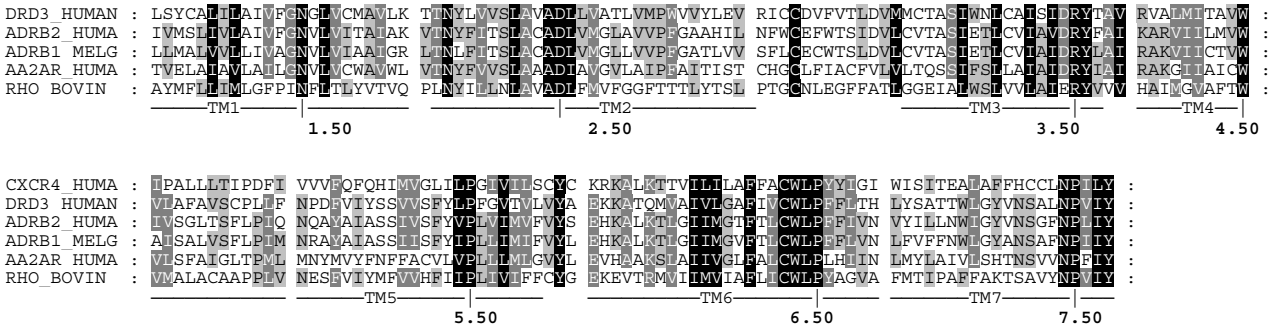
The human sequences of the D3 and CXCR4 receptors were used for homology modelling. The D3 and CXCR4 sequences as well as the sequences of the available templates, i.e. rhodopsin; the human beta 2 adrenergic, turkey beta 1 adrenergic and human A2A adenosine receptors were added to the sequence alignment of the transmembrane helices of 372 human G protein-coupled receptors and then aligned using ClustalX 1.81, Blosum 80 matrix (1). The conserved CYS of the second extracellular loop in the sequences of the investigated receptors and templates was then realigned manually. The final alignment used for homology modelling is shown bellow. The highest sequence identity and similarity between the investigated receptors and templates were identified with the beta 2 adrenergic receptor (42% and 66% for the D3 receptor, and 26% and 52% for the CXCR4). The crystal structure of the beta 2 adrenergic receptor with the PDB code - 2RH1 was employed to model the D3 and CXCR4 receptors using Molecular Operating Environment (MOE) software (2) with a default homology modeling protocol. Among ten generated models in both modelling campaigns the model with the best scoring was chosen for the 500 steps of minimization and the following docking studies.

Ligand Docking

The docking of eticlopride into the D3 model and IT1t into the CXCR4 model was conducted using the Schrodinger Induced-Fit protocol (3-5). In this protocol, the ligand is firstly docked to the rigid putative binding site using the Glide module of Schrodinger (3), then residue side chain repacking of the binding site is performed using the Prime module (5) and finally, the ligand is re-docked with Glide (3). The Glide grid box was centered on the $D110^{3.32}$ ($^{3.32}$ - Ballesteros and Weinstein residue nomenclature) in the D3 model and on the $Y116^{3.32}$, $D262^{6.58}$, and $E288^{7.39}$ in the CXCR4 model. These residues play a notable role in the ligand binding according to mutagenesis data (6-7). The Glide re-docking was conducted with the SP precision scoring function. The Induced-fit protocol with default settings were used for van der Waals scaling. Twenty docking poses were generated and five best solutions were selected taking into account mutagenesis data (6-11) and the scoring function values. The docking of CVX15 was performed manually using the molecular dynamics constrained based on the mutagenesis studies. The low energy conformation of CVX15 was generated using the MacroModel (12) conformational search with the mixed torsional/low-mode sampling protocol and initially docked to the binding site using Glide (3). Several docking orientations were taken for the constrained simulations in which Arg-s of CVX15 were forced to interact with $D262^{6.58}$ or $E288^{7.39}$. Two final solutions were selected based on the energy function.

Receptor and ligand preparation

The receptor and ligands were prepared using the Protein Preparation Wizard and Lig-Preparation modules of the Schrodinger software and applying the OPLS2005 force fields (13). The low energy conformations of eticlopride and IT1t obtained from the MacroModel conformational search (12) were used for the docking studies.



1. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice. Nucl. Acids Res. 22:4673-4680.
2. (2009) MOE. [2009.10] Ed., Chemical Computing Group, Inc., Montreal, Quebec, Canada
3. (2009) Glide. Version 5 Ed., Schrödinger, LLC, New York, NY
4. (2009) Schrödinger Suite 2008 Induced Fit Docking Protocol, Glide. Version 5.0 Ed., Schrödinger, LLC, New York, NY
5. (2005) Prime. 1.7 Ed., Schrödinger, LLC, New York, NY
6. Mansour, A., Meng, F., Meador-Woodruff, J.H., Taylor, L.P., Civelli, O., Akil, H.(1992) Site-directed mutagenesis of the human dopamine D2 receptor. Eur J Pharmacol. 1;227(2):205-14.
7. Wong,R.S., Bodart, V., Metz, M., Labrecque, J., Bridger, G., Fricker, S.P. (2008) Comparison of the potential multiple binding modes of bicyclam, monocylam, and noncyclam small-molecule CXC chemokine receptor 4 inhibitors. Mol Pharmacol. 74(6):1485-95.
8. Woodward, R., Daniell, S.J., Strange, P.G., Naylor, L.H. (1994) Structural studies on D2 dopamine receptors: mutation of a histidine residue specifically affects the binding of a subgroup of substituted benzamide drugs. J Neurochem. 62(5):1664-9.
9. Sartania, N., Strange, P.G. (1999) Role of conserved serine residues in the interaction of agonists with D3 dopamine receptors. J Neurochem. 72(6):2621-4.
10. Malmberg, A., Nordvall, G., Johansson, A.M,, Mohell, N., Hacksell, U. (1994) Molecular basis for the binding of 2-aminotetralins to human dopamine D2A and D3 receptors. Mol Pharmacol. 46(2):299-312.
11. Alberts, G.L., Pregenzer, J.F, Im, W.B. (1998) Contributions of cysteine 114 of the human D3 dopamine receptor to ligand binding and sensitivity to external oxidizing agents. Br J Pharmacol. 125(4):705-10
12. (2008) MacroModel 9.6 Ed., Schrödinger, LLC, New York, NY
13. (2009) Maestro 8.5. Version 5.5 Ed., Schrödinger, LLC, New York, NY

3713/7425

Irina Pogozheva (irinap@umich.edu), Andrei Lomize (almz@umich.edu)

Department of Medicinal Chemistry, College of Pharmacy, University of Michigan

Information and data used for modeling

Sequences of human CXCR4 (P61073) and human D3 receptor (P35462) were obtained from the UniProtKB/Swiss-Prot database. Alignments of sequences of CXCR4 with bovine rhodopsin (from PDB ID: 1u19) and of D3 receptor with turkey β1-adrenoreceptor (from PDB ID: 2vt4) were taken from GPCRDB and were subsequently corrected during iterative structural refinement. Cyclic 16-amino-acid peptide was aligned with polyphemusin I (from PDB ID: 1rkk).

Modeling algorithm and tools

Homology models of human CXCR4 and human D3 receptors were developed using structural templates of bovine rhodopsin (1u19) and turkey β1-adrenoreceptor (2vt4), respectively. The models were generated as previously described[1,2]. The procedure included five steps. In the first step, initial models were developed by residue substitution using template-target sequence alignments. In the second step (loop modeling), the C-terminus of TM6 and N-terminus of TM7 of CXCR4 were extended by two and one helical turns, respectively; some loops were generated using different GPCR templates (1u19 and 2z73 for EL2, EL3 and N-terminus and 2vt4 for IL2). The long extracellular loop 2 (EL2) of CXCR4 was inserted between helices (as in 1u19) in order to provide the binding surface for the functionally active cross-linked dimer of chemokine SDF1/CXCL12 that interacts with receptor N-terminus (PDB ID: 1ko5[3]) linked to EL3 via a disulfide bond. Loops of D3 receptor were taken from the template, except EL2, which was modeled as a β-hairpin at residues 174-183. In the third step (structure refinement), distance geometry calculations by DIANA[4] were performed using template-derived structural restraints to refine loop conformations, remove steric overlaps, and maximize the number of H-bonds between receptor residues with similar conservation patterns in multiple sequence alignments. In the fourth step (ligand docking), ligands were docked in cavities between helices in accordance with polarity matching. No experimental restraints were used. Eticlopride was docked in D3 receptor by similarity to ligands of adrenergic receptors (from PDB IDs: 2rh1, 2vt4). Low-energy conformations of small-molecules, isothiourea compound ("1t") and eticlopride were generated by QUANTA (Accelrys). The CXCR4-specific 16-amino-acid peptide was generated from the NMR structure of polyphemusin I (1rkk). In the fifth step, energy minimization of ligand-receptor complexes was performed using the CHARMm potentials (QUANTA, Accelrys) with ε 10 and the adopted basis Newton-Raphson minimization method (100 iterations).

Model ranking criteria are: (1) satisfaction of structural restraints during distance geometry calculations (DIANA score); (2) maximization of intramolecular H-bonds; (3) polarity matching of residues.

References:

1. Chai B.X., Pogozheva I.D., Lai Y.M., Li J.Y., Neubig R.R., Mosberg H.I., Gantz I. Receptor-antagonist interactions in the complexes of agouti and agouti-related protein with human melanocortin 1 and 4 receptors. *Biochemistry*, 2005, 44:3418-3431.
2. Pogozheva I.D., Chai B.X., Lomize A.L., Fong T.M., Weinberg D.H., Nargund R.P., Mulholland M.W., Gantz I., Mosberg H.I. Interactions of human melanocortin 4 receptor with nonpeptide and peptide agonists. *Biochemistry*, 2005, 44:11329-113341.

3. Veldkamp C.T., Seibert C., Peterson F.C., De la Cruz N.B., Haugner J.C. 3rd, Basnet H., Sakmar T.P., Volkman B.F. Structural basis of CXCR4 sulfotyrosine recognition by the chemokine SDF-1/CXCL12. *Sci Signal*., 2008, 1:ra4.
4. Güntert P.,Wüthrich K. Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *J. Biomol. NMR,* 1991,1:447-456.

3801

Nathan E. Hall

Drug Discovery Biology
Monash Institute for Pharmaceutical Sciences
Monash University
Nathan.Hall@monash.edu

GPCDOCK2010 D3R modelling

Information and data used for modeling:
The D3R sequence was manually aligned to the beta2 adrenoreceptor (pdb:2rh1). A deletion was accommodated in TM2, before D3R:P167 and selected regions in the loops were unaligned to accommodate different loop lengths. All calculations were carried out using the ICM suite of programs.[1,2,3]

Initial Homology Model:
16 homology models of the D3R were made using 2rh1 as a template following ICM's biased probability Monte-Carlo (BPMC) homology modelling protocol. The model with the lowest energy was chosen for further calculations to sample docking poses of eticlopride.

Eticlopride Binding Poses:
Two independent sets BPMC simulations were carried out with the two protonation enantiomers of eticlopride in the D3R binding site. For this, an 11 Å sphere round the eticlopride binding site was sampled with fully flexible sidechains and a fully flexible E2 loop. The remaining receptor was represented by multiple grid maps. Sixteen independent runs were carried with each of the two eticlopride protonation enantiomers. A 4.5 Å distance restraint between the protonated eticlopride nitrogen and Asp110/od2 was used to ensure eticlopride remained in the binding site. Six different eticlopride poses were selected based on lowest energy criteria.

Building Homology models around 6 different eticlopride poses
For each of the 6 eticlopride poses (representing both protonation enantiomers), multiple homology models were built with the same ICM method but with the addition of the eticlopride ligand tethered to the one of the six poses. In total 186 models were made, incorporating the different poses and variations in which loop residues are tethered to the beta2 template. Reduction of the tether strength during the PBMC simulation allows subtle movements of the ligand.

Model Validation and Selection
All models were subjected to VLS of DrugBank database which contains 53 compounds annotated as binding one of the dopamine receptors or are listed in PDSP KiDB as binding dopamine D3 receptor. All 186 models were ranked on the basis of enriching for dopamine antagonists and redocking ICM binding score for eticlopride. The 5 top models were chosen on the basis that they both enriched for dopamine ligands and scored highly in eticlopride redocking.

1. Abagyan, R., Totrov, M. & Kuznetsov, D. Icm: A New Method For Protein Modeling and Design: Applications To Docking and Structure Prediction From The Distorted Native Conformation. *J Comp Chem* **15**, 488-506 (1994).
2. Cardozo, T., Totrov, M. & Abagyan, R. Homology modeling by the ICM method. *Proteins* **23**, 403-14 (1995).
3. Abagyan, R. & Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol* **235**, 983-1002 (1994).

4374/8141

Molecular Modeling of Dopamine D3 receptor-Eticlopride Complex

Muhammad Muddassar,[1,2,3] Yang Zhang,[2] Ae Nim Pae[3], Jooyoung Lee[1]

[1]School of Computational Sciences, Korea Institute for Advanced Study, 130-722, Seoul, Republic of Korea
[2]Center for Computational Medicine and Bioinformatics, University of Michigan, 100 Washtenaw Avenue, Ann Arbor, MI 48109-2218, USA.
[3]Center for Neuro Medicine, Korea Institute of Science and Technology, 130-650, Seoul, Republic of Korea

Receptor structure prediction

The human dopamine D3 receptor amino acid sequence was threaded using LOMETS (1) threading programs, which identified other known GPCR structures in PDB (IDs: 2RH1, 3EML, 2VT4, 1U19 and 2Z73) as possible templates. Continuous fragments were excised from these templates and were used to assemble full-length models using a modified replica-exchange Monte Carlo simulation (2) where loop regions were constructed by *ab initio* modeling. The simulation decoys were clustered using SPICKER (3) and the cluster centroids was used as the next round of I-TASSER reassembly. The structures with the lowest energy in the second round of assembly simulation were selected and full-atomic models were refined using FG-MD (4). The receptor modeling procedure was fully automated.

Ligand-receptor docking

The molecular docking of eticlopride in the predicted I-TASSER models was performed using Glide software as implemented in the schrödinger package. The ligand was sketched and its ionic state was adjusted at physiological pH using the ligprep module. During the docking procedure the receptor structure was kept rigid, while the eticlopride molecule was allowed to be fully flexible. The docking box was centered on Asp 110 (5) and the residues located within 15Å from this amino acid were defined as the possible binding site during the docking simulations. Glide docking runs were performed using its “standard precision” mode, with a VdW (Van der Waals scaling) factor of 1.0 for non polar receptor atoms. These were defined as the atoms with partial charge lower than 0.25 for the receptor and lower than 0.15 for the ligand. Thirty eticlopride conformers were generated near Asp110 and out of which 10 poses were selected for further post docking minimization. The binding poses were selected based on the ionic interactions made by docked conformation of eticlopride with the Asp110, Glide score and visual inspection.

References

1. Wu ST, Zhang Y (2007) LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research* 35:3375-3382.
2. Wu S, Skolnick J, Zhang Y (2007) Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biol* 5:17.
3. Zhang Y, Skolnick J (2004) SPICKER: A clustering approach to identify near-native protein folds. *Journal of Computational Chemistry* 25:865-871.
4. Zhang J, Zhang Y (2010) High-resolution protein structure refinement using fragment guided molecular dynamics.Submitted.

5. Sokoloff P, Giros B, Martres MP, Bouthenet ML, Schwartz JC (1990) Molecular-Cloning and Characterization of a Novel Dopamine Receptor (D3) as a Target for Neuroleptics. *Nature* 347:146-151.

1. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols 5: 725-738.

2. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, et al. (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. Journal of Medicinal Chemistry 47: 1739-1749.

5084

Laura Lopez, Cristian Obiol-Pardo and Jana Selent

Computer Assisted Drug Design
Pompeu Fabra University (UPF/PRBB), Barcelona, Spain
jana.selent@upf.edu

Structural prediction of the D3 dopaminergic receptor in complex with eticlopride

Homology modeling. The human sequence of the dopamine D3 receptor was retrieved from the Swiss-Prot database. ClustalX software was used to align the sequence with the crystal structure of the β2 adrenergic G-protein-coupled receptor (GPCR) (PDB entry 2RH1). The resulting alignment was then manually refined to ensure perfect alignment of the highly conserved residues of the GPCR superfamily. The structural models of the receptors were built using the MODELLER suite of programs (1), which yielded 15 candidate models. The best structures were selected from these candidates, according to the MODELLER objective function and visual inspection. The resulting receptor structures were optimized by the AMBER99 force field (2) using the molecular modeling program MOE (Molecular Operating Environment; Chemical Computing Group, Inc). PROCHECK software (3) was used to assess the stereochemical quality of the minimized structures, resulting in good quality parameters and an excellent distribution of Ψ and Φ angles in the Ramachandran plot (more than 90 % of the residues are in the most favored regions).

Docking. Docking of eticlopride was carried out using the GOLD software (4), defining a centroid point in the residue Asp3.32 and expanding it 20 Å around this residue. One hundred genetic algorithms runs were submitted and further scored employing the ASP scoring function. The ligand was restricted to form a salt bridge between the positively charged nitrogen and the carboxylate of $Asp^{3.32}$. After clustering, best five scored solutions were retained and submitted to the conformational space analysis described below.

Conformational space analysis. The conformational space of the molecular systems (12 Å around the ligand) of the five best solutions was explored using the Low Mode Search which is a short molecular dynamics simulation using velocities with little kinetic energy on the high-frequency vibrational modes (default settings, MOE Software). The best solution for the five selected molecular systems was employed to further refinement.

Refinement. The refinement of the molecular systems consists in six steps:

(1) *Extracellular loop 2 (ECL2) relaxation*: MD simulation over 100 ps in which the complete protein as well as the ligand were kept fixed while the ECL2 was kept flexible using the MOE software (AMBER99 force field, Born solvation, 300 K, time step 2 fs).

(2) *Relaxation of the bindings site*: MD simulation over 100 ps in which the binding site (10 Å around the ligand), and the ECL2 were kept flexible while the ligand and protein backbone atoms were kept fixed (AMBER99 force field, Born solvation, 300 K, time step 2 fs).

(3) *Relaxation of the ligand in the binding site*: MD simulation over 100 ps in which the ligand, the binding site (10 Å around the ligand), and the ECL2 were kept flexible while the protein backbone atoms were kept fixed (MMFF94 force field (5), Born solvation, 300 K, time step 2 fs).

(4) *Searching for essential water molecules in the binding site*: The ligand was solvated around 5 Å using the MOE software.

(5) *Final MD simulation*: MD simulation over 100 ps in which the ligand, the binding site (10 Å around the ligand), and the ECL2 were kept flexible while protein backbone atoms were kept fixed (MMFF94 force field, Born solvation, 300 K, time step 2 fs)

(6) *Energy minimization of the final complex*: the best solutions were energy-minimized by applying gradient minimization until the RMS gradient was lower than 0.001 kcal $mol^{-1}Å^{-1}$.

Rescoring and Ranking. After MD refinement, the best three structures of the ligand were re-scored employing the GOLD software with the scoring functions: Gold-Score, ChemScore, ASP and PLP. After consensus scoring, last ranking was performed by taking into account the formation of two intramolecular hydrogen bonds in eticlopride.

References

1. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology. 1993; 234:779-815.
2. Cornell WD, Cieplak P, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc. 1995; 117:5179-5197.
3. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 1993;26:283-291.
4. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins. 2003; 52:609-623.
5. Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. Journal of Computational Chemistry. 1996; 17:490-519.

5207

Comparative modeling and docking studies of chemokine receptor CXCR4

Sadia Mahboob, Tim Werner and W. Bret Church

Group in Biomolecular Structure and Informatics, Faculty of Pharmacy, University of Sydney
bret.church@sydney.edu.au

The comparative models of the CXCR4 were developed considering the other high resolution GPCR structures as templates alone. Automated alignments only served as approximations and upon inspection were manually adjusted in order to ensure the alignment of highly conserved residues, to avoid insertions or deletions in the transmembrane domains (TM) and to confirm the conserved disulfide bridges between the Cys residues (Cys 28 and Cys 274 located in the N terminus and helix VII and Cys109 and Cys186 residing in the TM3 and the second extracellular loop).

Based on the highest sequence identity and similarity and for the purpose of keeping some diversity we chose the two ligand containing human β2 adrenergic (2RH1)[1] and human adenosine A_{2A} (3EML)[2] receptors as templates. The final alignments were submitted to Modeller (Discovery Studio v1.7 and standalone implementation, Modeller 9v8)[3-5] for the homology modeling. We created some diversity in the models based on the inclusion of various waters in the binding site and whether the original ligands were retained from the templates, and over thirty receptor models were created. The resulting models for the CXCR4 were assessed by conformational energy, steric clashes, observations of the disposition of key residues in the general GPCR architecture and close visual inspection of the extracellular loops with respect to the binding site and the residues placed in the vicinity of the binding cavity. The objective model assessments were performed by MOLPROBITY[6,7].

The small molecule ligand IT1t was then docked into the top ranked twenty models using GLIDE 5.5[8] and Induced Fit Docking[9]. The ligand poses were evaluated on the basis of their docking scores, occurrence of measured contacts, H bonds and ligand geometry preference. Generally our procedures were for the purpose of achieving a plausible binding pose, rather than the best overall comparative models. Our five models were chosen to reflect some diversity of the solutions we found among the top 40, with some bias toward satisfying hydrogen bonding.

References

1. Cherezov, V. et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science 318, 1258-1265 (2007).
2. Jaakola, V.P. et al. The 2.6 Ångstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. Science 322, 1211-1217 (2008).
3. Discovery Studio Modeling v. 1.7; Accelrys Inc, San Diego, USA (2007).
4. Eswar, N. et al. Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics Chapter 5, Unit 5.6 (2006).
5. Sali, A. & Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234, 779-815 (1993).
6. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66, 12-21 (2010).
7. Davis, I.W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35, W375-W383 (2007).
8. Glide v 5.5, Schrödinger, LLC, New York, NY (2009).

9. Sherman, W., Day, T., Jacobson, M.P., Friesner, R.A. & Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem 49, 534-553 (2006).

5334

Low-resolution modeling of protein-ligand complexes

Michal Brylinski, Tadashi Ando, Aysam Guerler, Hongyi Zhou and Jeffrey Skolnick*

Center for the Study of Systems Biology, Georgia Institute of Technology, Atlanta.

*corresponding author

G-protein coupled receptors (GPCRs) are a large protein family of transmembrane receptors that mediate a variety of biological effects through pathways involving the activation of G-proteins. The numerous physiological processes controlled by GPCRs make them particularly important targets in the treatment of many diseases. In the GPCR-dock experiment, we have applied a collection of tools to model the complex structures of CXCR4 and DRD3: TASSER^{mfret}, an extension of TASSER designed specifically for the modeling of membrane proteins, FINDSITE, a template-based method for ligand-binding pocket detection (1) and Q-Dock, a low-resolution flexible ligand docking approach that employs statistical potentials (2).

GPCR structure modeling

The models of both receptors were generated by TASSER^{mfret}. We used the latest version of the SP^3 threading algorithm (3) to identify the templates and to construct the target-template alignments. The identified templates of rhodopsin were excluded from the significant hits and only two known GPCR structures (2rh1_A and 2vt4_A) were included in the subsequent refinement by TASSER^{mfret}. The rhodopsin templates were excluded because their loops are buried inside the binding pocket. Finally, all-atom models were constructed from the TASSER^{mfret} generated models, which have only Cα coordinates and a significant number of potential clashes. Here, we built the ideal main-chain models as close as possible to the Cα traces and we optimized hydrogen bond formation by combining the RMSD score to the Cα traces and an H-bond score. The side-chain coordinates were copied from templates if the aligned target residue type is identical to that of the template. The rest of the side-chains were rebuilt by an in-house procedure that iteratively constructs the side-chain geometries one by one followed by optimization using the DFIRE (4) energy function.

Small molecule docking

Here, we applied our Ligand Homology Modeling approach to predict the binding poses of small ligand molecules in the binding pocket of the modeled GPCR structures, CXCR4 and DRD3. First, the available crystal structures of GPCRs complexed with small molecules were superimposed onto the modeled structure of the target receptors. Next, we used FINDSITE (1) to predict binding residues from the superimposed template structures. Subsequently, we employed FINDSITE^{LHM} (5) to extract binding ligands, which were used as ligand templates in similarity-based docking of the target compounds. The initial binding poses were finally refined using Q-Dock (2) with the aligned ligand substructures restrained to the corresponding substructures of the template-bound compounds. The resulting models were ranked by the total docking energy and submitted after visual inspection.

Peptide docking

We used the modeled CXCR4 structure to carry out protein structure alignments with the GIS-server (GANGSTA+ Internet Services) (6) against the ASTRAL40 (SCOP version 1.75 (7)) set of non-redundant protein structures. This way, we identified rhodopsin (PDB code 1U19) as the protein with the highest degree of structural similarity to our query structure (239 aligned residues at 2.74Å).

Analysis of the CXCR4/rhodopsin structure alignment revealed that the latter contains a loop segment (starting at residue 176-ARG) in close proximity to the putative peptide-binding region of CXCR4 predicted with FINDSITE (1). The second residue of this segment is 177-TYR, which forms hydrophobic interactions with one of the transmembrane helices. We used the coordinates of the identified loop to manually thread the target peptide sequence. After threading, we verified that the two cysteine residues of the target sequence are in close proximity such that they may establish a disulfide bridge. We repeated the same procedure using different rhodopsin structures (PDB codes 1F88, 1HZX, 1L9H) to construct several models for the binding peptide.

Next, the modeled complex structures were subject to energy minimization using AMBER (8). For natural amino acids, the ff03 force field (9) was used. The topology file for D-proline (dPro) was constructed by modifying the proline (Pro) topology file, while keeping the charges of dPro atoms assigned to the same value as in Pro. For L-1-naphthylalanine (Nal) and L-citrulline (Cit), topology files and AM1-BCC atomic charges were built by Antechamber. For charge estimation, the extended conformations of Ace-Nal-Nme and Ace-Cit-Nme were used, where Ace and Nme represent acetyl and N-methyl groups, respectively. For each model, steepest descent minimization for 1000 cycles and then conjugate gradient minimization for 9000 cycles were performed in vacuum with a cutoff distance of 15.0 Å for non-bonding interactions and a dielectric constant of 4. During minimization, all Cα atoms of GPCR were restrained using a harmonic potential with the force constant of 10 kcal/mol/$Å^2$. Final models were ranked by the total energy.

1) Brylinski M, Skolnick J (2008) *PNAS* **105**:129-34.
2) Brylinski M, Skolnick J (2010) *J Comput Chem* **31**:1093-105.
3) Zhou H, Zhou Y (2005) *Proteins* **58**:321-8.
4) Zhou H, Zhou Y (2002) *Protein Sci* **11**:2714-26.
5) Brylinski M, Skolnick J (2009) *PLoS Comput Biol* **5**:e1000405.
6) Guerler A, Knapp EW (2010) *Nucl Acid Res* **38**:W46-52.
7) Hubbard TJ, Murzin AG, Brenner SE, Chothia C (1997) *Nucl Acid Res* **25**:236-9.
8) Case DA, Cheatham TE, Simmerling CL, Wang J, Duke, RE, Luo R, Wang B, Pearlman, DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA (2004) AMBER 8.
9) Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) *J Comput Chem* **24**:1999-2012.

5508

Henri Xhaard

Centre for Drug Research, Faculty of Pharmacy, University of Helsinki, Helsinki, Finland
henri.xhaard@helsinki.fi

Modeling of the CXCR4 receptor in complex with a cyclic peptide.

Information and data used for modeling

The single model presented was constructed based on the following starting hypotheses:

- from my previous experience with cyclic peptides binding to serine proteases (for example, PDB codes 1OX1, 1SFI and 1SMF), the peptide could be cyclic and interact with CXCR4 as a beta-hairpin that forms a beta-sheet with a loop in the receptor.
- The peptide does not reach the orthosteric site. This is influenced by the solution structure of SDF1 in complex with the receptor N-terminus (PDB code 2K03)
- The P8-P9 region of the peptide is the site of the beta-turn in the peptide
- The extracellular loop 2 forms a beta-hairpin that interacts with the peptide. The residues D181 and D182 forms the turn in the loop
- Excess material in extracellular loop 3 forms an alpha-helix
- The Nterminus is disordered (freely assigned by Modeller)

No experimental data was used (little time allocated to the study)

Modeling algorithms and tools

A simple way to build such docking mode was to take advantage of the loop conformation seen in bovine rhodopsin (PDB code 1U19). The model presented is a kind of "structural mimicry" where the cyclic peptide replaces the sheet 1 and 2 and is stacked on EC2 forming a 4-strands beta sheet. Because the peptide was hypothesized to bind to the loop region, the TM region of the template did not matter. Rhodopsin bovine (1U19) was used for homology modeling. Sybyl X (Tripos Ldt) was used as a molecular modeling platform. Modeller 9v2 (Sali and Blundel, 1993) was used to build ten initial models. Side chains interactions, in particular ionic interactions with the peptide, were manually optimized and modeller re-run to optimize side chains (some side-chains were deleted from template to explore flexibility). After adding the modified amino acids an energy minimization (100 steps) was performed. The peptide was not fully cyclized and normal prolines were used (little time allocated to the study) No specific ranking was done (single model).

Modeling of the CXCR4 receptor in complex with small molecule ligand

The peptide was removed from the model above and the small molecule manually docked at three different locations/orientations of the orthosteric site. An energy minimization (100 steps) was performed. The final binding modes are not fully relaxed. No specific ranking was done (little time allocated to the study).

Modeling of the D3 receptor in complex with small molecule ligand

Information and data used for modeling

- The D3 is evolutionary the closest to the beta adrenoceptors so the beta2-adrenoceptor is the best template (PDB code 1RH1)(see e.g. Xhaard et al., J.Med.Chem 2006)
- The conformation of eticlopride was studied using the Cambridge Structural Database (CSD code DOJXEF10). Internal hydrogen bonds were noted.

- Eticlopride should bind in its protonated; form ion-pair with D3.32
- Eticlopride binding mode should resemble the binding modes of ligands (including carazolol) to the beta-2 adrenoceptor (2RH1)
- The Nterminus is disordered (freely assigned by Modeller)

No experimental data was otherwise used (little time allocated to the study)

Modeling algorithms and tools

A model of D3 was derived using Modeller 9v2. Carazolol bound to the beta-2 structural template (superimposed on the D3 model) was used to guide manual docking in Sybyl-X. Side chains conformations were manually modified. One hundred steps of energy minimization. Steric clashes remained close to the amine of the ligand (little time allocated to the study). No specific ranking was done (a single model that looked very reasonable was suggested).

6006

Wiktor Jurkowski and Arne Elofsson

Center of Biomembrane Research
Stockholm University
wiktor.jurkowski@cbr.su.se

Modeling of receptors

Protein sequences were extracted from provided template files. PDB entries of structural templates and all other information used for structure prediction and docking are listed in the table below.

The pcons.net server (Wallner 2007) was used for automated prediction of the receptors structures. In each run it produces a ranked list of models based on the templated of various homologous protein templates. All downloaded models were further ranked by the coverage of target sequence and trans-membrane topology. An alignment was assumed to be correct when it contained a minimal number of gaps in the TM regions. Finally, the core structures of both CXCR4 and D3 was predicted based on β-2-Adrenergic receptor (2RH1) (Cherezov 2007).

C-X-C chemokine receptor type 4
Three variants of disulphide bridge are possible on receptors extracellular side, between cysteins 109, 186 and 274. The most probable one is a bond between Cys109 and Cys186, which resembles the type of bond existing in adrenergic or adenosine receptors in opposite to rhodopsin which represents the remaining two variants. Here, the grid for molecular docking simulations was centred on Asn106, Lys282 and Phe174.

Dopamine D3 Receptor
Theoretically 6 combinations of disulphide bonds can be generated among the four cystein residues in extracellular loops. We only studied combination which would not close the ligand binding site. Therefore, we excluded presence of bonds between ECL2 and ECL3 and let the connection to be created between Cys252-Cys255 within ECL3 and between ECL1 and ECL2, Cys103-Cys181. Similar arrangement of disulphide bonds is present in the structure of adenosine receptor.

We built the structure of extracellular loops ECL1-ECL3 using Modeller 9v8 (Melo 2007). 100 models of each loop were clustered and ranked by Modeller and ProQ (Wallner 2003) scores. Loops were constructed with disulphide bridge restraints and underwent the same scoring procedure as core structures. Conformations were additionally rewarded for opening of receptor TM binding site. At the end 10 automatically selected models were passed to docking stage.

Ligand docking and selection

In most docking programs the conformation of the ligand is searched and tested on rigid receptor. This can lead to many conformations which internal energy is only roughly estimated. To filter preferable ligand geometries we applied classification procedure which combines Macro Model (Schrödinger, Inc) conformational search and DFT/B3LYP geometry optimisation in Jaguar (Schrödinger, Inc). Further, generated poses are clustered and geometry is optimized for each cluster representative. All low energy structures creates starting points for screening and additionally serve as checkpoints in

docking to favour predefined geometries of docked poses. Ligand libraries were docked to receptors using Glide (Schrödinger, Inc) and Vina (Trott 2010). Consensus ranking made by scoring functions of both programs enabled us to primarily select three receptor models in which complexes are ranked highest despite of ligand conformation. We finally picked the best ligand poses after clustering and filtering of favoured geometries. No information about experimentally derived protein-ligand contacts were used in our study.

We applied this docking procedure to target 1 and 3. The peptide ligand of target 2 was too big to treat with the same degree of flexibility hence another approach was more suitable. Here, ligand preparation differs by the fact that cluster representatives were optimized in a classic alforcefield. Peptide library was docked with Zdock (Mintseris 2007) with restraints to search only on the extra-cellular receptor site. 18 selected clusters representing 54000 poses generated on a 6A grid was selected. Finalluy, complexes were refined using Macro Model by steepest descent and conjugent gradient minimisations.

References:

Cherezov, V., *et al.*, High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318,** 1258-65 (2007).

Melo, F., Sali, A., 2007. Fold assessment for comparative protein structure modeling. Protein Sci 16, 2412-2426.

Mintseris et al. Integrating statistical pair potentials into protein complex prediction. Proteins (2007) vol. 69 (3) pp. 511-20

Schrödinger, Inc.: Portland, OR, 1991-2000.

Trott and Olson. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. (2010) vol. 31 (2) pp. 455-61

Wallner and Elofsson. Can correct protein models be identified?. Protein Sci (2003) vol. 12 (5) pp. 1073-86

Wallner, B., Larsson, P. and Elofsson, A. (2007) Pcons.net: protein structure prediction meta server. *Nucleic Acids Res* **35** (suppl_2) : W369-W374.

7141

Methods for Generating Homology Models of D_3 Receptor and Docking of Eticlopride

Ahsan K. Murad, Malgorzata Drwal, Tom B. Dupree, Renate Griffith*

School of Medical Sciences, University of New South Wales, Sydney, Australia.
*: corresponding author. Email: r.griffith@unsw.edu.au

Template selection: Sequence similarities for transmembrane helices (TMH) were evaluated with CLUSTAL-W (http://www.uniprot.org/). 2RH1 and 2VT4 were the best templates.
Multiple sequence alignment: Using CLUSTAL-W, the following sequences were aligned: human adrenoceptors (ARs); α_{1A}, α_{1B}, α_{1D}, β_2; common turkey β_1 AR; bovine and squid rhodopsin; human adenosine receptor 2A and dopamine D_3 receptor.
Model generation: Modeler, as implemented in Accelrys Discovery Studio 2.5 (DS2.5) (http://accelrys.com/) was used for models 1-4. Default parameters were used, except for "cut overhangs", which was set to "no". The best model according to Modeler was chosen. SwissModel in alignment mode (http://swissmodel.expasy.org/) was used for model 5. The pairwise alignments were hand-edited to conform with the multiple sequence alignment, and also to ensure maximum conserved residues and minimum gaps in the helices and at the end of TMH5 up to residue 217, and formation of the disulfide bond between C103 and C181. The L119W mutation was made. Chain B of 2VT4 was used. The lysozyme sequence was deleted from 2RH1. Residues 218 to 320 were deleted from the models. The quality of models was ssessed using Protein Health Report in DS2.5 (Ramachandran assessment).
Structure-based pharmacophores: To propose binding modes for eticlopride, all features were mapped with DS2.5 for the ligands in the two templates. Features not corresponding to observed interactions in the crystal structures were deleted. A conformational model was prepared for eticlopride (using the 'best' option), and fitted to the pharmacophores. The best-fitting conformation was used after checking its relative energy.
Docking: Gold 4.1 (http://www.ccdc.cam.ac.uk/products/gold_suite/) was used with Goldscore. D110 was defined as an interaction motif. Default parameters were used, except for 100 docking runs per ligand, population size 1000, maximum number of genetic algorithm operations $1x10^6$. Crystal structure ligands were docked back into the crystal structures as a control. Two conformations of protonated eticlopride were used: one was sketched and 'cleaned' in DS2.5, one fitted to the structure-based pharmacophore. The binding pocket was chosen based on crystal structures (residues within 7 Å around ligand), then the cavity detection algorithm was used. Flexible sidechains were defined based on residues in the crystal structure that interact with the ligand: F106, V107, D110, V111, I183, S192, F346, H349, T369, Y373. For each run, the best pose was chosen according to a number of criteria: number and size of clusters (RMSD of 2.0 Å, using heavy atoms, compared to each other), number of hydrogen bond interactions, particulary with D110, best scoring pose in chosen cluster, comparison to ligand in crystal structure.
Final model selection: Based on Procheck (http://deposit.pdb.org/validate/), appearance of ECL2 and effect on docked poses, overall quality of TMHs by visual inspection, hydrogen bond interactions, docking scores, similarity to ligand poses in crystal structures.
Construction of chimeric proteins: To add N-termini, all models were aligned to a model based on bovine rhodopsin (1U19) using Cα atoms of 5 conserved residues in the middle of TMHs as tethers (V50, D110, S165, S196, and I333). The Cα of residues 30-34 were then aligned for each model with the rhodopsin based model, the best aligned peptide bond was chosen, and all rhodopsin-based residues downstream and all model residues upstream deleted. A new peptide bond was created. DS2.5 models

based on 2VT4 had a problem at the end of TMH5. This was replaced as described for the N-termini. Atom numbering and order and other problems of DS2.5 pdb files were corrected by saving models as mol2 files, opening in pymol (http://www.pymol.org/) and saving as pdb files in pymol.

7334

Liliana Ostopovici-Halip[1], Cristian Bologa[2]

[1]Institute of Chemistry, Romanian Academy, Timisoara, Romania
[2]Division of Biocomputing, University of New Mexico
lili.ostopovici@gmail.com
cbologa@salud.unm.edu

An identical working protocol was used for both human dopamine D3 and CXCR4 receptors.

1)Homology modeling

Crystal structure of the β2-adrenergic receptor (β2-AR) with bound carazolol (PDB ID: 2rh1) [1]was used as template structure for homology modeling of the human D3 and CXCR4 receptors. The sequence alignments were initially generated using the 3D-Jury metaserver [2]. They were then manually adjusted in order to avoid deletions or insertions in the transmembrane domains and to preserve the highly-conserved amino acid motifs specific for each transmembrane helix identified based on the conserved residues within the GPCR amino acid sequences. The Modeller package (version 9v8)[3, 4] was used to generate the three-dimensional models for the human D3 and CXCR4 receptors based on the X-ray structure of β2-AR using the sequence alignment presented in figure 1. Five models were generated for each sequence target. The resulted models were first refined in order to reduce the side chain sterical clashes and then the whole receptors, including the main backbone were minimized. When necessary, the second extracellular loop (EL2) was built without any alignment between target and template sequences using the Modeller's routine MODLOOP.

2)Ligand docking

The molecular docking of the ligands was performed with Glide [5] implemented in the Schrödinger package. The ligands were prepared for docking using LigPrep 2.1 application [6] as follows: the hydrogen atoms were added, the possible tautomers at physiological conditions were generated, and the ionization states were set. Finally, the final geometries were minimized using the OPLS_2005 force field.

The receptors have been prepared using the Protein Preparation Workflow [7]. All the hydrogen atoms were added to the homology models, the charge state of Asp, Glu, Arg Lys and His residues and the orientation of hydroxyl and amide groups were optimized. At the end a restrained minimization was performed by the Impref utility, which is based on the Impact molecular mechanics engine [49] and the OPLS2005 force field, setting a max RMSD of 0.20.

Docking protocol. The grid-box was centered on the 3.33 aminoacid position and the size of the box was set at 30Å on each side. The final poses are ranked by a receptor-ligand interaction energy (Glide score).

3) Criteria for prediction analysis and ranking.

The resulted poses have been clustered based on shape similarity and the most representative five conformations have been selected for each ligand-receptor pair. Ranking has been chosen based on the Glide score of each cluster.

References

1. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human beta2-

adrenergic G protein-coupled receptor. Science. 2007, 318(5854):1258-65
2. Ginalski K, Elofsson A, Fischer D, Rychlewski L. "3D-Jury: a simple approach to improve protein structure predictions." Bioinformatics. 2003 May 22;19(8):1015-8
3. Eswar, N. et al. Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Suppl 15, 5.6.1-5.6.30, 200
4. Marti-Renom, M.A. et al. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325 (2000).
5. Glide 45; Schrodinger LLC, New York, NY 2008.
6. LigPrep version 21; Schrodinger, LLC, New York, NY 2007.
7. Protein Preparation Wizard; Schrodinger LLC, New York, NY 2007.

```
2RH1   : ---------------------------------DEVWVVGMGIVMSLIVLAIVFGNVLVITAIAKFE
hD3    : ---------MASLSQLSSHLNYTCGAENSTGASQARPHAYYALSYCALILAIVFGNGLVCMAVLKER
hCXCR4: MEGISIYTSDNYTEEMGSGDYDSMKEPCFREENANFNKIFLPTIYSIIFLTGIVGNGLVILVMGYQK

2RH1   : RLQTVTNYFITSLACADLVMGLAVVPFGAAHILMK-MWTFGNFWCEFWTSIDVLCVTASIETLCVIA
hD3    : ALQTTTNYLVVSLAVADLLVATLVMPWVVYLEVTGGVWNFSRICCDVFVTLDVMMCTASILNLCAIS
hCXCR4: KLRSMTDKYRLHLSVADLLFVITLPFWAVDAVAN---WYFGNFLCKAVHVIYTVNLYSSVLILAFIS

2RH1   : VDRYFAITSPFKY---QSLLTKNKARVIILMVWIVSGLTSFLPIQMHWYRATHQEAINCYAEETCCD
hD3    : IDRYTAVVMPVHYQHGTGQSSCRRVALMITAVWVLAFAVSCPLLFGFNTT---------GDPTVCSI
hCXCR4: LDRYLAIVHATNS---QRPRKLLAEKVVYVGVWIPALLLTIPDFIFANVS--------EADDRYICD

2RH1   : FFTNQ-----AYAIASSIVSFYVPLVIMVFVYSRVFQEAKRQLNIFEMLRIDEGLRLKIYKDTEGYY
hD3    : SNPD-------FVIYSSVVSFYLPFGVTVLVYARIYVVLKQRRRKR--------------------
hCXCR4: RFYPNDLWVVVFQFQHIMVGLILPGIVILSCYCIIISKLSHSK-----------------------

2RH1   : TIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITKDEAEKLFNQDVDAAVRGILRNAKLKPVYDSLD
hD3    : ----------------------------------------ILTRQNSQCNSVRPGFPQQTLSPDPA
hCXCR4: ------------------------------------------------------------------

2RH1   : AVRRAALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYNQTPNRAKRVITTFRTGTWD
hD3    : HLELKRYYSICQDTALGGPGFQERGGELKREEKTRNSLSPTIAPKLSLEVRKLSNGRLSTSLKLGPL
hCXCR4: ------------------------------------------------------------------

2RH1   : AY-KFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIRK----------EVYILLNWIGY
hD3    : QPRGVPLREKKATQMVAIVLGAFIVCWLPFFLTHVLNTHCQTCHVS---------PELYSATTWLGY
hCXCR4: ------HQKRKALKTTVILILAFFACWLPYYIGISIDSFILLEIIKQGCEFENTVHKWISITEALAF

2RH1   : VNSGFNPLIYCRSP-DFRIAFQELLCL--------------------------------
hD3    : VNSALNPVIYTTFNIEFRKAFLKILSC--------------------------------
hCXCR4: FHCCLNPILYAFLGAKFKTSAQHALTSVSRGSSLKILSKGKRGGHSSVSTESESSSFHSS
```

Figure 1. Sequence alignment for the hD3, hCXCR4 and β2-adrenergic receptors

7533

Chen, K.M.[1], Sun, J.[2], Barth, P.[1,2]

[1] Verna and Marrs Mclean Department of Biochemistry and Molecular Biology
[2] Department of Pharmacology, Baylor College of Medicine, Houston, TX, USA.
patrickb@bcm.edu

Information and data used for modeling. The CXCR4 and D2DR receptor structures were modeled using the beta2 adrenergic receptor structure (2RH1) as a homolog template. Sequence/template alignments were performed using the HHpred server. No experimental data were used other than the template structure.

Modeling algorithms and criteria for scoring. The receptor structures were modeled using a comparative modeling technique developed for membrane protein structures (1). The regions of the target sequences poorly aligned with the template were modeled *de novo* by Monte-Carlo-based peptide fragment insertion. The well-aligned regions of the query sequence were threaded onto the structural template. When the poorly aligned regions were refolded, the entire structure was subjected to all-atom refinement using an energy function developed for membrane protein structures (1, 2). Around 20000 independent simulations were performed for each target sequence. For the receptor alone models, the 10% lowest energy receptor structures were clustered based on their helical structure. For the receptor with docked ligand models, the 10% lowest energy receptor structures were clustered based on the putative ligand-binding regions and the centers of the largest clusters were used for docking the ligand. Ligand docking uses a combination of Monte Carlo moves, sidechain repacking, and gradient-based minimization with the Rosetta all-atom potential. Receptor backbone and sidechain torsions along with ligand torsions, position, and orientation were optimized simultaneously (3). The ligand conformer libraries were generated using Omega (OpenEye Inc.). 5000 docking trajectories were run for each receptor model and the 5% lowest-energy models were selected and ranked based on the binding energy with the ligand. The models with the lowest binding energy were inspected visually and selected based on chemical intuition.

References

1. Chen et al., unpublished results.
2. Barth et al., 2007 PNAS, 104(40), 15682-15687
3. Davis & Baker, 2009 JMB, 385(2):381-92.

7571

Vladimir Yarov-Yarovoy and David Baker

University of Washington
yarovoy@uw.edu, dabaker@uw.edu

Information and data used for modeling

Multiple homologous templates for modeling of human CXCR4 and D3 G-protein coupled receptors (GPCR) were identified using HHPred homology detection server1-3. Top ranking templates included: human β2-adrenergic G protein-coupled receptor (pdb: 2RH1), bovine rhodopsin (pdb: 1U19), β1 adrenergic receptor (pdb: 2VT4), squid rhodopsin (pdb: 2Z73), and human A2A Adenosine Receptor (pdb: 3EML). No additional experimental data were used during modeling. Sequence alignments between human CXCR4 and D3 and the templates were generated using HHPred homology detection server [1-3].

Modeling algorithms and tools

Rosetta loop building protocol [4] and a modified version of Rosetta-Membrane full atom energy function [5] (Yarov-Yarovoy V. and Baker D., unpublished data) were used to generate loop conformations of CXCR4 and D3 GPCRs. Modified version of the Rosetta-Membrane full-atom method incorporated the following changes to energy function and protocol: (1) Coarse-grained membrane environment energy term was added to the energy function and derived from statistical analysis of non-homologous set of membrane protein structures [6] (Yarov-Yarovoy V. and Baker D., unpublished data); (2) Relative weights of hydrogen bonding energies between side chain –side chain and side chain – backbone atoms were increased by a factor of 2; (3) Membrane normal and center search were performed only before the full atom conformational search. 5,000 models were generated for each starting template using the Rosetta loop building protocol and ranked by total Rosetta-Membrane full-atom energy. Notably, all lowest energy models originated from template based on human β2-adrenergic G protein-coupled receptor (pdb: 2RH1). The lowest energy conformations from the Rosetta loop building protocol were then used as input to Rosetta-Membrane full atom refinement protocol5 where backbone torsion angles and side chain rotamer conformations were optimized using Monte Carlo minimization procedure and a modified version of Rosetta-Membrane full atom energy (see above). No ligand docking was performed. 4,000 models were generated and five lowest energy models were submitted as the best models.

References

1. Soding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951-60 (2005).
2. Soding, J., Biegert, A. & Lupas, A.N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244-8 (2005).
3. Hildebrand, A., Remmert, M., Biegert, A. & Soding, J. Fast and accurate automatic structure prediction with HHpred. *Proteins* **77 Suppl 9**, 128-32 (2009).
4. Wang, C., Bradley, P. & Baker, D. Protein-protein docking with backbone flexibility. *J Mol Biol* **373**, 503-19 (2007).
5. Barth, P., Schonbrun, J. & Baker, D. Toward high-resolution prediction and design of transmembrane helical protein structures. *Proc Natl Acad Sci U S A* **104**, 15682-7 (2007).
6. Yarov-Yarovoy, V., Schonbrun, J. & Baker, D. Multipass membrane protein structure prediction using Rosetta. *Proteins* **62**, 1010-1025 (2006).

8004

Bas Vroling, Marijn P.A. Sanders, Jan Klomp & Sander B. Nabuurs*

Computational Drug Discovery Group, Centre for Molecular and Biomolecular Informatics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

E-mail: s.nabuurs@cmbi.ru.nl

Homology modeling

DRD3: PDB entry 2RH1 was used as a template to build a homology model. The structure was cleaned, the lysozyme protein was removed and the bound ligands retained. The residues in the loop between TM5 and TM6 (residues 218-317) were discarded in the modeling process. Modeling was performed using the Yasara program and its built-in modeling algorithm[1]. The generated model was manually refined and a final energy minimization step was applied.

CXCR4: A hybrid template was created and used to build a homology model. PDB entry 2RH1 was used as the base template. Residues in the loop between TM6 and TM7 (residues 267-276) were discarded in the modeling process. The loop between TM4-TM5 was deleted (residue Y174-A198) and replaced with the loop TM4-TM5 from PDB entry 3EML (AA2AR_HUMAN, residues N144-N175). Further homology modeling as described above.

Pharmacophore searching

Structure-based pharmacophores based on the transmembrane domains of the constructed homology models were built using the Snooker program[2] (http://www.cmbi.ru.nl/software/snooker). For both targets default settings were used, except for the procedure to identify the ligand binding pocket in CXCR4. Here, parameters were slightly modified to generate a ligand binding pocket in between the transmembrane helices.

The Snooker method prioritizes residue positions based on sequence conservation in a multiple sequence alignment and deduces the key interacting residues from the derived statistics. Known DRD3 and CXCR4 actives were retrieved from Chembl (https://www.ebi.ac.uk/chembl) using a 50nM activity cutoff and 3D conformations were generated using Cyndi[3]. A pharmacophore search was performed to identify the structure based pharmacophore for each target complementary to most of the active compounds. 3D conformations of both eticlopride and 1t were generated using Cyndi and the ligands were matched in their corresponding structure based pharmacophores. All pharmacophore matching poses were used for high resolution pose generation.

Docking

The low-resolution binding modes obtained from Snooker were used to guide the high-resolution molecular docking using the Fleksy program[4] (http://www.cmbi.ru.nl/software/fleksy). The various orientations of eticlopride in DRD3 and 1t in CXCR4 were used as anchors to guide induced fit docking using the Fleksy protocol. The generated poses were optimized in the respective homology models and subsequently ranked using a consensus scoring function utilizing the FlexX and PLP docking scores, geometrical quality indicators and molecular dynamics force field interaction energies.

1. Krieger, E.; Joo, K.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K., Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins 2009, 77 Suppl 9, 114-22.
2. Sanders, M.P.A.; Verhoeven, S.; de Graaf, C.; Roumen, L.; de Vlieg, J.; Klomp, J.P.G., Snooker: A structure based pharmacophore generation tool applied to class A GPCRs, manuscript in preparation.
3. Liu, X.; Bai, F.; Ouyang, S.; Wang, X.; Li, H.; Jiang, H., Cyndi: a multi-objective evolution algorithm based method for bioactive molecular conformational generation. BMC Bioinformatics 2009, 10, 101.
4. Nabuurs, S. B.; Wagener, M.; de Vlieg, J., A flexible approach to induced fit docking. Journal of Medicinal Chemistry 2007, 50, (26), 6507-18.

8241

Gregory V. Nikiforovich

MolLife Design LLC, 751 Aramis Drive, St. Louis, MO 63141
gnikiforovich@gmail.com

TM region of CXCR4 *(modeling procedure as in [1])***.** Boundaries of TM helices were determined by sequential alignment to rhodopsin by Clustal-W. Individual helices were subjected to energy minimization with some restraints and were aligned to the structural template 1F88 (rhodopsin).
Extracellular loops of CXCR4 *(modeling procedure as in [2])***.** Various possible options for conformations of the EC loops were obtained completely *de novo* with the assumption of the conservative disulfide bridge C109-C186. They were restored on the rigid template obtained for the TM region. The "most open" conformations of the EC loops were selected for further considerations.
N-term 28-41 of CXCR4. All low-energy backbone conformations without sterical clashes with TM region and EC loops were obtained by build-up modeling procedure (as in [1]).
N-term + EC loops of CXCR4. The specific "most open" EC loop conformation allowing closing of the C28-C274 disulfide bridge was selected as a template for docking of possible CVX conformations.
CVX. Low-energy backbone conformations of CVX were obtained by build-up procedure. Only one low-energy cluster of conformations for the 4-13 cycle was obtained.

Orienting conformations of CVX in the N-term + EC loops template.
1) ***According to orientations of CPP predicted earlier***. Earlier, we have predicted two possible orientations of cyclo(*Tyr-Arg-Arg-Nal-Gly*), CPP [3]. Obtained conformations of CVX were oriented in the EC loop template by fitting spatial arrangement of the side chains Arg^{1}-Arg^{2}-Nal^{3} and Tyr^{5} to *Arg-Arg-Nal* and *Tyr*, respectively (first set of poses), or the side chains of Arg^{1}-Arg^{2}-Nal^{3} to *Arg-Arg-Nal* and Arg^{14} to Arg^{2} (second set of poses).
2) According to de novo GRAMM procedure. CVX conformation with the closest ends was docked to the N-term + EC loops template by the GRAMM procedure (downloaded from http://vakser.bioinformatics.ku.edu/main/resources_gramm.php). Out of best poses, those with contacts to CXCR4 residues implicated as functionally important were selected.

Final energy minimization and selection of models. All selected poses of CVX were incorporated in the CXCR4 template (N-term + EC loops + TM region) and subjected to energy minimization that included closing of the two disulfide bridges, C109-C186 and C28-C274. Five final models of CXCR4-CVX complex were selected by the following criteria: *(i)* low energy of the complexes; *(ii)* most perspective contacts of Arg's with the negative charged residues of CXCR4; *(iii)* close contacts of Arg's, Tyr and Nal to functionally important CXCR4 residues.

References

[1] Nikiforovich GV, Marshall GR, Baranski TJ. Modeling Molecular Mechanisms of Binding of the Anaphylotoxin C5a to the C5a Receptor. Biochemistry 2008;47:3117-3130.
[2] Nikiforovich GV, Taylor CM, Marshall GR, Baranski TJ. Modeling the possible conformations of the extracellular loops in G-protein-coupled receptors. Proteins: Structure, Function, and Genetics 2010;78:271-285.
[3] Vabeno J, Nikiforovich GV, Marshall GR. Insight into the binding mode for cyclopentapeptide antagonists of the CXCR4 receptor. Chemical Biology & Drug Design 2006;67(5):346-354.