Supplementary Information for:

Enhancing structure prediction and design of soluble and membrane proteins with explicit solvent-protein interactions

Jason K. Lai [1,§], Joaquin Ambia [1,†,§], Yumeng Wang [3], & Patrick Barth [1,2,3,¶,&]

[1] Department of Pharmacology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.

[2] Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.

[3] Structural and Computational Biology and Molecular Biophysics Graduate Program, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.

[†] Present address: Center for Petroleum and Geosystems Engineering, Cockrell School of Engineering, The University of Texas at Austin, Austin, TX 78712, USA.

[§] The authors contributed equally to the study.

[¶] Correspondence: P.B. (patrickb@bcm.edu)

[&] Present address: Interfaculty Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Federale de Lausanne, CH-1015 Lausanne, Switzerland

Figure S1

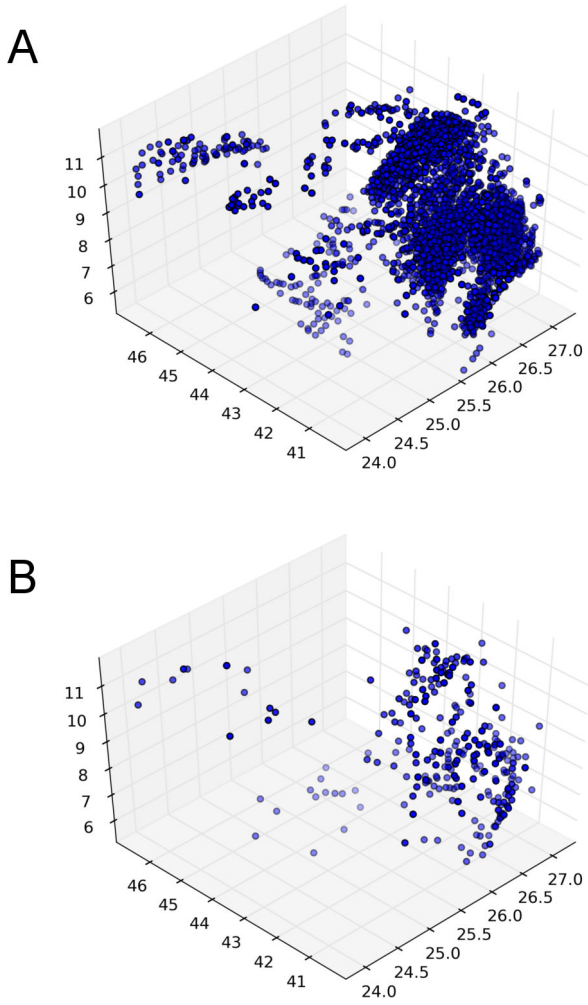

Figure S1. Distribution of a water molecule rotamer set on an XYZ coordinate plot after a 30-fold size reduction, Related to STAR Methods text. The generated water rotamer sets are shown (A) before and (B) after the rotamer filtering process (**methods**). The space coverage of the pre-filtered rotamer set contains nearly 15,000 rotamers (A) while the post-filtered rotamer set contains 500 (B).

Figure S2

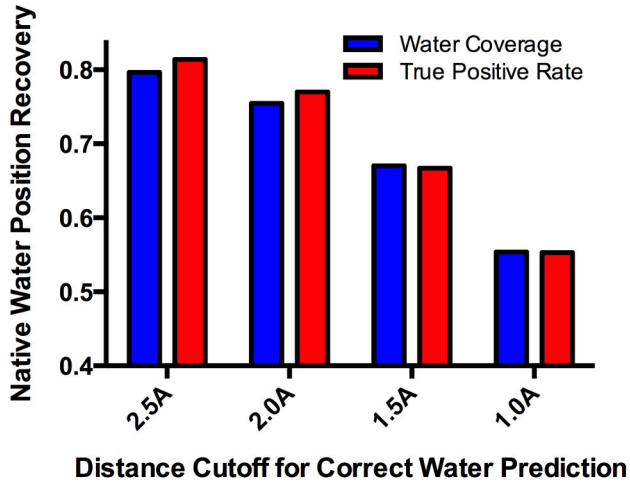

Figure S2. SPaDES prediction of protein-bound waters positions with varying levels of stringency, Related to Figure 2. (A) Recovery of experimentally-observed water molecules (blue) and true positive rates of predicted water molecules (red) are shown at distance cutoff of 2.5 Å, 2.0 Å, 1.5 Å, and 1.0 Å between predicted and experimentally-resolved waters. Higher fractional rates, ranging from zero to one, indicate increased accuracy of the predictions.

Figure S3

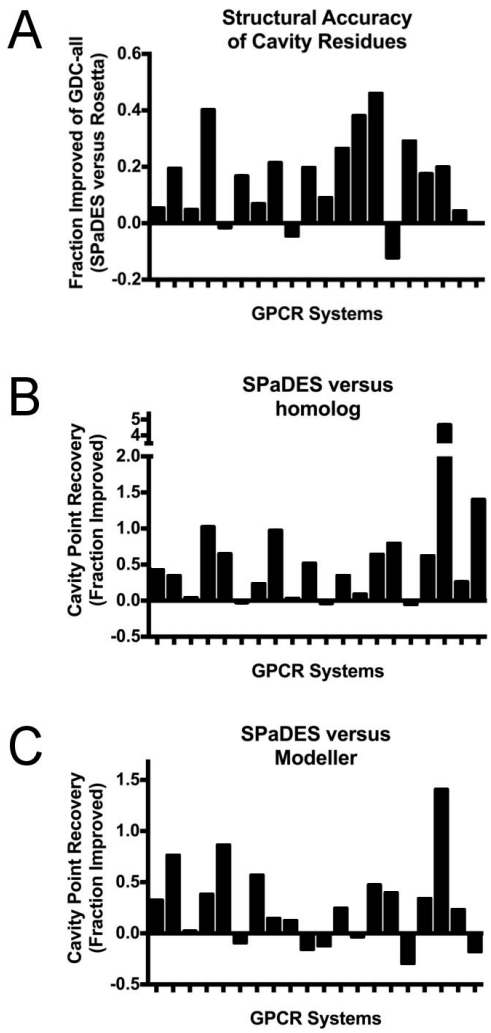

Figure S3. High-resolution prediction of solvated transmembrane regions in G protein-coupled receptor homology models, Related to Figure 5. (A) Improved all-atom structure prediction of local regions around buried cavities (3 Å) (measured by fraction improved of GDC-all) using SPaDES. Comparison between models generated using RosettaMembrane implemented with the hybrid solvation model of SPaDES and RosettaMembrane implemented with an implicit solvation model. (B-C) Improved structure prediction of buried cavities in GPCRs modeled using RosettaMembrane implemented with the hybrid solvation model of SPaDES compared to the starting homolog templates (B) and compared to homology models generated by Modeller (C). The structures of twenty GPCRs were modeled starting from distant homologs (sequence identity between 20 and 40%) as described in **Figure 5**. Recovery of cavity lattice points was used to determine the fraction improvement of hybrid solvation predictions over template and Modeller.

Figure S4

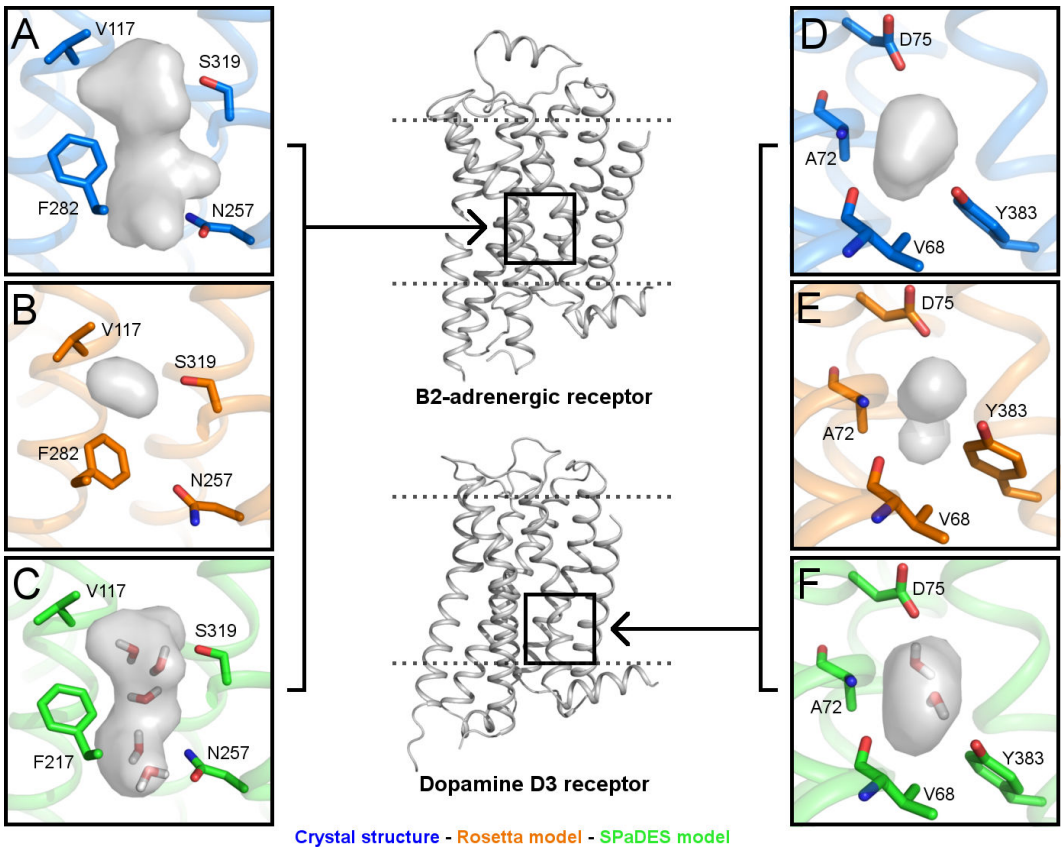

Figure S4. Accurate prediction of hydrated cavities in GPCRs using SPaDES, Related to Figure 5. (A-C) Beta 2 adrenergic receptor modeled from the adenosine A2a receptor. (D-F) Dopamine D3 receptor modeled from the beta 2 adrenergic receptor. These cavities are mostly lost when modeled using implicit solvation (B, E). The cavities are shown as gray surfaces, neighboring side-chains that affect the cavity shape are shown in sticks with text labels, and predicted explicit waters when using the hybrid scoring function are shown as red sticks in (C, F). Panels have been rotated to show all marked components. X-ray structures (A, D); SPaDES models (C, F).

Figure S5

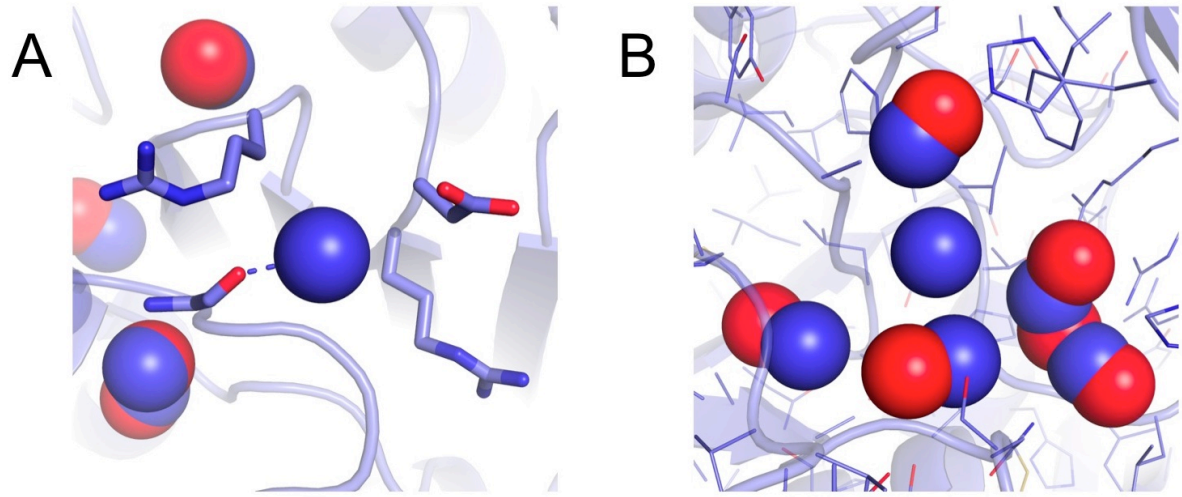

Figure S5. SPaDES limitations in predicting the position of water molecules, Related to STAR Methods text. (A) In the X-ray structure of the SHV-1 beta-lactamase-BLIP complex, an experimentally-observed water molecule forming only one weak polar interaction with a neighboring protein backbone carbonyl group was not recapitulated by SPaDES. (B) In the X-ray structure of the Subtilisin Calsberg-Eglin C complex, an experimentally-observed water molecule in the center of a large cavity and coordinated by surrounding water molecules was not recapitulated by SPaDES. Experimentally-observed and predicted water molecules by SPaDES are shown as blue and red spheres, respectively.

Table S1. Water molecule position recovery at protein-protein binding interfaces using SPaDES, Fold-X with explicit water molecule modeling, and HADDOCK refinement, Related to Figure 2. Columns 4, 5. Number of water molecules experimentally detected or predicted at the binding interface. Columns 6, 7. Water cov. (water coverage): Fraction of native water molecules recovered by a predicted water molecule within a given distance threshold (top performing of the 5 lowest energy models, standard error of mean). Columns 8, 9. T.P. (true positive): Fraction of predicted water molecules within a given distance threshold of a native water molecule (top performing of the 5 lowest energy model, standard error of mean). Standard error of mean (std. err.) is calculated from the 5 lowest energy models. *A2AR unfolded during the HADDOCK molecular dynamics refinement and thus was not included in the analysis for HADDOCK.

	Protein-Protein Complex			SPaDES: 2.5Å Distance Threshold				
PDB	**Protein 1**	**Protein 2**	**# Exp. waters**	**# Pred. waters**	**Water cov. rate**	**Water cov. std. err.**	**T.P. water rate**	**T.P. water std. err.**
1ACB	Bovine alpha-chymotrypsin	Eglin C	16	24	0.94	0.01	0.58	0.02
1BRS	Barnase	Barstar	21	22	0.86	0.01	0.95	0.00
1CSE	Subtilisin Carlsberg	Eglin C	26	17	0.62	0.01	0.94	0.00
1DQJ	HyHEL-63 Fab	HEW Lysozyme	18	24	0.72	0.03	0.54	0.02
1DVF	IgG1-kappa D1.3 Fv	E5.2 Fv	8	16	0.63	0.00	0.44	0.02
1JTG	TEM-1 beta-latamase	BLIP	38	39	0.82	0.02	0.90	0.02
1TM1	Subtilisin BPN	Chymotrypsin inhibitor 2	23	24	0.78	0.02	0.88	0.03
1XD3	UCH-L3	Ubiquitin	32	26	0.75	0.02	0.88	0.01
2FTL	Bovine trypsin	BPTI	22	20	0.82	0.02	0.85	0.01
2G2U	SHV-1 beta-lactamase	BLIP	47	50	0.83	0.03	0.88	0.02
2WPT	Colicin E2 immunity protein	Colicin E9 Dnase	10	12	1.00	0.00	0.83	0.02
3NPS	Membrane-type serine protease 1	S4 Fab	15	12	0.60	0.01	0.75	0.02
4EIY	A2AAR		35	31	0.80	0.01	0.97	0.01
1C3W	Bacteriorhodopsin		9	9	1.00	0.02	1.00	0.03
		Averages:			0.80	0.02	0.81	0.02

	Protein-Protein Complex			SPaDES: 2Å Distance Threshold				
PDB	**Protein 1**	**Protein 2**	**# Exp. waters**	**# Pred. waters**	**Water cov. rate**	**Water cov. std. err.**	**T.P. water rate**	**T.P. water std. err.**
1ACB	Bovine alpha-chymotrypsin	Eglin C	16	24	0.88	0.02	0.54	0.02
1BRS	Barnase	Barstar	21	22	0.86	0.01	0.95	0.00
1CSE	Subtilisin Carlsberg	Eglin C	26	17	0.58	0.00	0.94	0.01
1DQJ	HyHEL-63 Fab	HEW Lysozyme	18	24	0.61	0.03	0.46	0.02
1DVF	IgG1-kappa D1.3 Fv	E5.2 Fv	8	16	0.63	0.00	0.44	0.01
1JTG	TEM-1 beta-latamase	BLIP	38	39	0.82	0.02	0.79	0.02
1TM1	Subtilisin BPN	Chymotrypsin inhibitor 2	23	24	0.74	0.02	0.83	0.02
1XD3	UCH-L3	Ubiquitin	32	26	0.75	0.02	0.88	0.01
2FTL	Bovine trypsin	BPTI	22	20	0.77	0.01	0.85	0.01
2G2U	SHV-1 beta-lactamase	BLIP	47	50	0.72	0.02	0.76	0.02
2WPT	Colicin E2 immunity protein	Colicin E9 Dnase	10	12	1.00	0.00	0.83	0.02
3NPS	Membrane-type serine protease 1	S4 Fab	15	12	0.53	0.01	0.67	0.02
4EIY	A2AAR		35	31	0.80	0.01	0.94	0.01
1C3W	Bacteriorhodopsin		9	9	0.89	0.03	0.89	0.03
		Averages:			0.75	0.01	0.77	0.02

	Protein-Protein Complex			SPaDES: 1.5Å Distance Threshold				
PDB	**Protein 1**	**Protein 2**	**# Exp. waters**	**# Pred. waters**	**Water cov. rate**	**Water cov. std. err.**	**T.P. water rate**	**T.P. water std. err.**
1ACB	Bovine alpha-chymotrypsin	Eglin C	16	24	0.81	0.03	0.54	0.02
1BRS	Barnase	Barstar	21	22	0.67	0.03	0.68	0.04
1CSE	Subtilisin Carlsberg	Eglin C	26	17	0.58	0.01	0.88	0.03
1DQJ	HyHEL-63 Fab	HEW Lysozyme	18	24	0.39	0.03	0.29	0.03
1DVF	IgG1-kappa D1.3 Fv	E5.2 Fv	8	16	0.38	0.05	0.19	0.03
1JTG	TEM-1 beta-latamase	BLIP	38	39	0.74	0.01	0.69	0.01
1TM1	Subtilisin BPN	Chymotrypsin inhibitor 2	23	24	0.70	0.03	0.67	0.02
1XD3	UCH-L3	Ubiquitin	32	26	0.69	0.01	0.85	0.02
2FTL	Bovine trypsin	BPTI	22	20	0.77	0.02	0.85	0.02
2G2U	SHV-1 beta-lactamase	BLIP	47	50	0.70	0.03	0.66	0.02
2WPT	Colicin E2 immunity protein	Colicin E9 Dnase	10	12	1.00	0.02	0.83	0.03
3NPS	Membrane-type serine protease 1	S4 Fab	15	12	0.53	0.01	0.67	0.02
4EIY	A2AAR		35	31	0.77	0.01	0.87	0.01
1C3W	Bacteriorhodopsin		9	9	0.67	0.03	0.67	0.03
		Averages:			0.67	0.02	0.67	0.02

	Protein-Protein Complex			SPaDes: 1Å Distance Threshold				
PDB	**Protein 1**	**Protein 2**	**# Exp. waters**	**# Pred. waters**	**Water cov. rate**	**Water cov. std. err.**	**T.P. water rate**	**T.P. water std. err.**
1ACB	Bovine alpha-chymotrypsin	Eglin C	16	24	0.69	0.03	0.46	0.02
1BRS	Barnase	Barstar	21	22	0.48	0.04	0.45	0.05
1CSE	Subtilisin Carlsberg	Eglin C	26	17	0.46	0.01	0.71	0.03
1DQJ	HyHEL-63 Fab	HEW Lysozyme	18	24	0.28	0.02	0.21	0.02
1DVF	IgG1-kappa D1.3 Fv	E5.2 Fv	8	16	0.25	0.05	0.13	0.03
1JTG	TEM-1 beta-latamase	BLIP	38	39	0.58	0.03	0.56	0.02
1TM1	Subtilisin BPN	Chymotrypsin inhibitor 2	23	24	0.52	0.01	0.50	0.03
1XD3	UCH-L3	Ubiquitin	32	26	0.53	0.01	0.65	0.01
2FTL	Bovine trypsin	BPTI	22	20	0.73	0.02	0.80	0.03
2G2U	SHV-1 beta-lactamase	BLIP	47	50	0.60	0.04	0.56	0.03
2WPT	Colicin E2 immunity protein	Colicin E9 Dnase	10	12	0.90	0.02	0.75	0.02
3NPS	Membrane-type serine protease 1	S4 Fab	15	12	0.53	0.02	0.67	0.02
4EIY	A2AAR		35	31	0.66	0.03	0.74	0.04
1C3W	Bacteriorhodopsin		9	9	0.56	0.04	0.56	0.04
		Averages:			0.55	0.03	0.55	0.03

	Protein-Protein Complex			Fold-X: 2Å Distance Threshold				
PDB	**Protein 1**	**Protein 2**	**# Exp. waters**	**# Pred. waters**	**Water cov. rate**	**Water cov. std. err.**	**T.P. water rate**	**T.P. water std. err.**
1ACB	Bovine alpha-chymotrypsin	Eglin C	16	41	0.81	-	0.39	-
1BRS	Barnase	Barstar	21	26	0.81	-	0.69	-
1CSE	Subtilisin Carlsberg	Eglin C	26	25	0.58	-	0.56	-
1DQJ	HyHEL-63 Fab	HEW Lysozyme	18	39	0.67	-	0.33	-
1DVF	IgG1-kappa D1.3 Fv	E5.2 Fv	8	28	0.50	-	0.14	-
1JTG	TEM-1 beta-latamase	BLIP	38	64	0.61	-	0.36	-
1TM1	Subtilisin BPN	Chymotrypsin inhibitor 2	23	37	0.70	-	0.51	-
1XD3	UCH-L3	Ubiquitin	32	44	0.56	-	0.48	-
2FTL	Bovine trypsin	BPTI	22	41	0.68	-	0.44	-
2G2U	SHV-1 beta-lactamase	BLIP	47	62	0.60	-	0.47	-
2WPT	Colicin E2 immunity protein	Colicin E9 Dnase	10	33	1.00	-	0.39	-
3NPS	Membrane-type serine protease 1	S4 Fab	15	29	0.67	-	0.34	-
4EIY	A2AAR		35	34	0.60	-	0.65	-
1C3W	Bacteriorhodopsin		9	16	0.78	-	0.44	-
		Averages:			0.67	-	0.44	-

	Protein-Protein Complex			HADDOCK: 2Å Distance Threshold				
PDB	**Protein 1**	**Protein 2**	**# Exp. waters**	**# Pred. waters**	**Water cov. rate**	**Water cov. std. err.**	**T.P. water rate**	**T.P. water std. err.**
1ACB	Bovine alpha-chymotrypsin	Eglin C	16	20	0.56	0.02	0.45	0.04
1BRS	Barnase	Barstar	21	14	0.52	0.01	0.79	0.02
1CSE	Subtilisin Carlsberg	Eglin C	26	14	0.38	0.00	0.71	0.03
1DQJ	HyHEL-63 Fab	HEW Lysozyme	18	20	0.44	0.03	0.40	0.04
1DVF	IgG1-kappa D1.3 Fv	E5.2 Fv	8	23	0.75	0.00	0.30	0.03
1JTG	TEM-1 beta-latamase	BLIP	38	36	0.53	0.02	0.53	0.04
1TM1	Subtilisin BPN	Chymotrypsin inhibitor 2	23	13	0.52	0.02	0.85	0.04
1XD3	UCH-L3	Ubiquitin	32	26	0.59	0.02	0.65	0.04
2FTL	Bovine trypsin	BPTI	22	19	0.36	0.02	0.42	0.03
2G2U	SHV-1 beta-lactamase	BLIP	47	48	0.51	0.02	0.46	0.04
2WPT	Colicin E2 immunity protein	Colicin E9 Dnase	10	9	0.50	0.00	0.56	0.04
3NPS	Membrane-type serine protease 1	S4 Fab	15	16	0.53	0.02	0.50	0.04
1C3W	Bacteriorhodopsin		9	4	0.33	0.03	0.75	0.05
		Averages:			0.50	0.02	0.57	0.04

Table S2. Side-chain rotamer recovery at protein-protein binding interfaces, Related to Figure 3. Recoveries are reported for Rosetta, SPaDES, Fold-X with implicit solvation and Fold-X with explicit water molecule modeling for two benchmarks: high-resolution hydrated and low- and high-resolution sets (**methods**). Standard error of mean (std. error) is calculated from the 5 lowest energy models for Rosetta and SPaDES.

High-Resolution Hydrated Set								
	Rosetta				**SPaDES**			
	# Correctly predicted rotamers	**# Total rotamers**	**Fraction of correctly predicted**	**Fraction Std. Error**	**# Correctly predicted rotamers**	**# Total rotamers**	**Fraction of correctly predicted**	**Fraction Std. Error**
Polar and charged sidechains	86	119	0.72	0.00	95	119	0.80	0.02
Hydrophobic sidechains	14	16	0.88	0.00	14	16	0.88	0.00
Complexes with <20 interface waters	31	49	0.63	0.00	34	49	0.69	0.01
Complexes with >20 interface waters	69	86	0.80	0.00	75	86	0.87	0.01
All benchmark sidechains	100	135	0.74	0.00	109	135	0.81	0.02
	Fold-X (implicit)				**Fold-X**			
	# Correctly predicted rotamers	**# Total rotamers**	**Fraction of correctly predicted**	**Fraction Std. Error**	**# Correctly predicted rotamers**	**# Total rotamers**	**Fraction of correctly predicted**	**Fraction Std. Error**
Polar and charged sidechains	68	119	0.57	-	78	119	0.66	-
Hydrophobic sidechains	14	16	0.88	-	14	16	0.88	-
Complexes with <20 interface waters	25	49	0.51	-	29	49	0.59	-
Complexes with >20 interface waters	57	86	0.66	-	63	86	0.73	-
All benchmark sidechains	82	135	0.61	-	92	135	0.68	-

Low- and High-Resolution Set								
	Rosetta				**SPaDES**			
	# Correctly predicted rotamers	**# Total rotamers**	**Fraction of correctly predicted**	**Fraction Std. Error**	**# Correctly predicted rotamers**	**# Total rotamers**	**Fraction of correctly predicted**	**Fraction Std. Error**
Polar and charged sidechains	237	352	0.67	0.01	262	352	0.74	0.04
Hydrophobic sidechains	44	50	0.88	0.00	44	50	0.88	0.00
Complexes with <20 interface waters	124	167	0.74	0.01	130	167	0.78	0.03
Complexes with >20 interface waters	157	235	0.67	0.01	176	235	0.75	0.02
All benchmark sidechains	281	402	0.70	0.01	306	402	0.76	0.03
	Fold-X (implicit)				**Fold-X**			
	# Correctly predicted rotamers	**# Total rotamers**	**Fraction of correctly predicted**	**Fraction Std. Error**	**# Correctly predicted rotamers**	**# Total rotamers**	**Fraction of correctly predicted**	**Fraction Std. Error**
Polar and charged sidechains	203	352	0.58	-	192	352	0.55	-
Hydrophobic sidechains	41	50	0.82	-	44	50	0.88	-
Complexes with <20 interface waters	105	167	0.63	-	107	167	0.64	-
Complexes with >20 interface waters	139	235	0.59	-	129	235	0.55	-
All benchmark sidechains	244	402	0.61	-	236	402	0.59	-

Table S3. Prediction of mutational effects on protein binding energies, Related to Figure 4. Predictions are reported for Rosetta, SPaDES, Fold-X with implicit solvation (Fold-X implicit) and Fold-X with explicit water molecule modeling (Fold-X) for two benchmarks: a high-resolution set of protein-protein complex structures with experimentally detected water molecules at the binding interface (high-resolution hydrated set) and a complete (low- and high-resolution) set of protein-protein complex structures (**methods**). *Fraction of predicted $\Delta\Delta G_{binding}$ that are correctly predicted to be either stabilizing (<-1 kcal.mol^{-1}), destabilizing (>1 kcal.mol^{-1}), or neutral (>-1 and <1 kcal.mol^{-1}).

Rosetta			
High-Resolution Hydrated Set		**Low- and High-Resolution Set**	
# of complexes	12	# of complexes	39
# of mutations	120	# of mutations	532
R correlation coefficient	0.56	R correlation coefficient	0.33
Standard error of fit	2.76	Standard error of fit	2.27
Slope of best fit	0.99	Slope of best fit	0.40
Stability classification*	0.59	Stability classification*	0.48
SPaDES			
High-Resolution Hydrated Set		**Low- and High-Resolution Set**	
# of complexes	12	# of complexes	39
# of mutations	120	# of mutations	532
R correlation coefficient	0.72	R correlation coefficient	0.60
Standard error of fit	2.32	Standard error of fit	1.93
Slope of best fit	0.43	Slope of best fit	0.31
Stability classification*	0.62	Stability classification*	0.55

Fold-X (implicit)			
High-Resolution Hydrated Set		**Low- and High-Resolution Set**	
# of complexes	12	# of complexes	39
# of mutations	120	# of mutations	532
R correlation coefficient	0.59	R correlation coefficient	0.48
Standard error of fit	2.69	Standard error of fit	2.10
Slope of best fit	1.11	Slope of best fit	0.81
Stability classification*	0.64	Stability classification*	0.59
Fold-X			
High-Resolution Hydrated Set		**Low- and High-Resolution Set**	
# of complexes	12	# of complexes	39
# of mutations	120	# of mutations	532
R correlation coefficient	0.47	R correlation coefficient	0.39
Standard error of fit	2.92	Standard error of fit	2.22
Slope of best fit	0.90	Slope of best fit	0.63
Stability classification*	0.60	Stability classification*	0.57

Table S4. Detailed energetic deconstruction of the prediction of mutational effects on protein binding energies using SPaDES, Related to Figure 4. $\Delta\Delta G_{binding}$ predictions are reported for Rosetta and SPaDES (**methods**) for 4 selected mutations described in **Figure 4**. Black text: specific interactions between water molecules and residues correspond to those highlighted in **Figure 4**. Red text: additional water-mediated interactions not displayed in **Figure 4**. $\Delta\Delta G_{binding}$ is reported in kcal/mol for both predicted and experimental values. Predicted values in Rosetta Energy Units were translated into kcal/mol using the slope of the best linear correlation fit between predicted and experimental values. The number of hydrogen bonds for each water molecule highlighted in **Figure 4** is reported for the wild type and mutated binding interfaces. Numbers in parentheses correspond to strong hydrogen bonds with Energy<-0.5 REU.

Colicin E2 immunity protein - Colicin E9 Dnase (S50A)

$\Delta\Delta G_{binding}$ Summary		Wildtype water hydrogen bonds			Mutant water hydrogen bonds			Wildtype hbond count		Mutant hbond count	
	$\Delta\Delta G_{binding}$ (kcal.mol^{-1}.K^{-1})	acceptor	donor	hbond energy (REU)	acceptor	donor	hbond energy (REU)	water molecule	number of hydrogen bonds	water molecule	number of hydrogen bonds
Experimental $\Delta\Delta G_{binding}$	2.42	wat1	S50	-1.13	wat2	S50A	-1.16	wat1	3 (3)	wat1	2 (2)
SPaDES $\Delta\Delta G_{binding}$	2.63	wat2	S50	-0.99	T87	wat1	-1.44	wat2	3 (3)	wat2	3 (2)
Rosetta $\Delta\Delta G_{binding}$	1.01	S50	wat2	-1.32	D51	wat1	-1.41				
SPaDES standard error	0.20	T87	wat1	-1.44	Q92	wat2	-0.95				
Rosetta standard error	1.42	D51	wat1	-1.33	E41	wat2	-0.22				
		Q92	wat2	-1.26							

Barnase-Barstar (T42A)

$\Delta\Delta G_{binding}$ Summary		Wildtype water hydrogen bonds			Mutant water hydrogen bonds			Wildtype hbond count		Mutant hbond count	
	$\Delta\Delta G_{binding}$ (kcal.mol^{-1}.K^{-1})	acceptor	donor	hbond energy (REU)	acceptor	donor	hbond energy (REU)	water molecule	number of hydrogen bonds	water molecule	number of hydrogen bonds
Experimental $\Delta\Delta G_{binding}$	2.35	T42	wat1	-1.21	D39	wat1	-1.18	wat1	4 (4)	wat1	3 (3)
SPaDES $\Delta\Delta G_{binding}$	1.56	wat1	R83	-0.68	wat2	wat1	-0.55	wat2	4 (3)	wat2	4 (3)
Rosetta $\Delta\Delta G_{binding}$	-0.37	wat1	wat2	-1.06	wat3	wat1	-1.05	wat3	4 (4)	wat3	4 (4)
SPaDES standard error	0.79	wat3	wat1	-0.98	wat3	wat2	-0.79				
Rosetta standard error	2.72	wat3	wat2	-0.59	D39	wat3	-1.18				
		D39	wat3	-1.47	wat2	R83	-0.38				
		wat2	R83	-0.08	wat4	wat2	-1.11				
		wat2	wat4	-0.57	E73	wat3	-1.12				
		E73	wat3	-1.29							

HyHEL-63 Fab-HEW Lysozyme (Y96A)

$\Delta\Delta G_{binding}$ Summary		Wildtype water hydrogen bonds			Mutant water hydrogen bonds			Wildtype hbond count		Mutant hbond count	
	$\Delta\Delta G_{binding}$ (kcal.mol^{-1}.K^{-1})	acceptor	donor	hbond energy (REU)	acceptor	donor	hbond energy (REU)	water molecule	number of hydrogen bonds	water molecule	number of hydrogen bonds
Experimental $\Delta\Delta G_{binding}$	1.13	wat1	Y96a	-0.77	S91	wat1	-1.32	wat1	4 (4)	wat1	4 (3)
SPaDES $\Delta\Delta G_{binding}$	1.37	wat1	R21	-1.27	wat2	wat1	-0.80	wat2	4 (4)	wat2	3 (3)
Rosetta $\Delta\Delta G_{binding}$	-2.70	wat2	wat3	-0.59	W98	wat4	-1.44	wat3	4 (4)	wat3	--
SPaDES standard error	0.24	wat3	wat4	-0.62	wat4	wat5	-1.09	wat4	3 (3)	wat4	3 (3)
Rosetta standard error	3.84	W98	wat4	-1.39	Y20	wat1	-1.02	wat5	3 (3)	wat5	2 (2)
		wat4	wat5	-1.00	wat2	S91	-0.63				
		S91	wat1	-1.11	Y50	wat2	-1.31				
		wat2	S91	-0.59	wat5	Q89	-0.94				
		Y20	wat1	-1.08	wat1	R21	-0.41				
		Y20	wat2	-1.22	Y47	wat4	-1.23				
		Y50	wat2	-1.22							
		Y50	wat3	-1.13							
		S91	wat3	-1.20							
		wat5	Q89	-1.28							
		wat6	wat5	-0.51							

Barnase-Barstar (E73A)

$\Delta\Delta G_{binding}$ Summary		Wildtype water hydrogen bonds			Mutant water hydrogen bonds			Wildtype hbond count		Mutant hbond count	
	$\Delta\Delta G_{binding}$ (kcal.mol^{-1}.K^{-1})	acceptor	donor	hbond energy (REU)	acceptor	donor	hbond energy (REU)	water molecule	number of hydrogen bonds	water molecule	number of hydrogen bonds
Experimental $\Delta\Delta G_{binding}$	2.35	E73	wat1	-1.20	wat1	wat6	-0.73	wat1	3 (2)	wat1	3 (3)
SPaDES $\Delta\Delta G_{binding}$	1.56	E73	wat2	-1.26	wat1	wat7	-0.68	wat2	4 (3)	wat2	4 (3)
Rosetta $\Delta\Delta G_{binding}$	-0.37	wat3	R83	-0.59	wat2	wat6	-1.08	wat3	5 (3)	wat3	3 (2)
SPaDES standard error	0.79	wat3	K27	-0.95	wat2	wat7	-1.26	wat4	3 (1)	wat4	3 (3)
Rosetta standard error	2.72	wat2	K27	-0.50	wat4	wat8	-0.54	wat5	3 (2)	wat5	4 (2)
		wat4	D54	-1.22	wat8	wat3	-0.82			wat6	4 (4)
		D75	wat5	-1.10	D39	wat3	-1.04			wat7	4 (4)
		wat5	wat3	-1.43	wat7	Y103	-0.72			wat8	3 (3)
		I55	wat2	-1.29	wat5	wat8	-0.74				
		wat1	wat9	-0.71	wat4	wat5	-1.13				
		Y103	wat1	-0.16	D54	wat4	-1.44				
		wat2	wat10	-0.26	I55	wat2	-1.31				
		E73	wat3	-0.28	wat6	K27	-0.50				
		D39	wat3	-0.12	wat6	wat7	-0.78				
		D54	wat4	-0.25	wat9	wat1	-1.28				
		wat4	G53	-0.03	wat10	wat2	-0.19				
		wat5	R83	-0.12	wat3	R83	-0.44				
					I51	wat5	-0.33				
					wat5	R83	-0.26				

Table S5. Improved homology modeling of GPCRs using SPaDES compared to Rosetta, Related to Figure 5. Sequence identity reported between selected GPCR target and homolog template for the entire TM region and for local regions around the buried TM cavities. Improved all-atom structure prediction (measured by fraction improved of GDC-all) and improved cavity geometry prediction (measured by fraction improved of recovered cavity points) of SPaDES versus Rosetta: The top performing of the 5 lowest energy structures and the standard error of mean (SEM) are provided. Standard error of mean is calculated from the 5 lowest energy models.

Target PDB	Target	Template PDB	Template	Seqence identity of trans-membrane region	Sequence identity of residues around cavity (3Å)	Sequence identity of residues around cavity (9Å)	GDC-all of cav. res. (Frac. improved)	GDC-all of cav. res. (Frac. improved SEM)	Cavity point recovery (Frac. improved)	Cavity point recovery (Frac. improved SEM)
3PBL	Dopamine D3 receptor	4BVN	Beta1-adenoceptor	39	50	41	0.054	0.019	0.111	0.126
4IAR	Chimeric 5-HT1B-BRIL	3RZE	Histamine H1 receptor	37	48	42	0.194	0.033	0.101	0.106
3PBL	Dopamine D3 receptor	2RH1	Beta2-adenergic receptor	36	49	39	0.048	0.017	0.395	0.115
3PBL	Dopamine D3 receptor	5CXV	M1 muscarinic acetylcholine recptor	34	43	37	0.402	0.024	0.249	0.103
4EIY	Chimeric A2a adenosine receptor	4BVN	Beta1-adenoceptor	33	44	38	-0.015	0.023	0.535	0.100
3EML	A2A adenosine receptor	2RH1	Beta2-adenergic receptor	33	42	35	0.167	0.020	0.192	0.087
4EIY	Chimeric A2a adenosine receptor	2RH1	Beta2-adenergic receptor	33	46	36	0.069	0.025	0.050	0.131
3PBL	Dopamine D3 receptor	5DSG	M4 muscarinic acetylcholine receptor	33	43	36	0.214	0.044	0.278	0.097
2RH1	Beta2-adenergic receptor	3EML	A2a adenosine receptor	32	43	38	-0.046	0.019	0.888	0.185
5CXV	M1 muscarinic acetylcholine recptor	4BVN	Beta1-adenoceptor	30	39	32	0.197	0.033	0.073	0.122
5DSG	M4 muscarinic acetylcholine receptor	4BVN	Beta1-adenoceptor	30	39	29	0.091	0.021	0.197	0.117
3PBL	Dopamine D3 receptor	4S0V	OX2 orexin receptor	29	40	30	0.265	0.025	0.271	0.086
3PBL	Dopamine D3 receptor	3EML	A2a adenosine receptor	29	39	34	0.381	0.036	0.254	0.176
3V2Y	Sphingosine 1-phosphate receptor-1	5CXV	M1 muscarinic acetylcholine recptor	29	35	29	0.461	0.048	0.244	0.125
3V2Y	Sphingosine 1-phosphate receptor-1	5DSG	M4 muscarinic acetylcholine receptor	26	31	29	-0.122	0.017	0.275	0.140
3PBL	Dopamine D3 receptor	3V2Y	Sphingosine 1-phosphate receptor-1	25	33	30	0.292	0.028	0.345	0.196
4S0V	OX2 orexin receptor	5DSG	M4 muscarinic acetylcholine receptor	22	33	22	0.175	0.022	0.352	0.092
3VW7	Protease-activated receptor 1	4S0V	OX2 orexin receptor	22	31	25	0.199	0.050	0.172	0.117
3PBL	Dopamine D3 receptor	1U19	Bovine rhodopsin	20	28	27	0.044	0.067	0.009	0.139
4XNV	P2Y1 receptor	5DSG	M4 muscarinic acetylcholine receptor	20	31	25	0.000	0.067	0.732	0.263

Table S6. Polar residues in representative GPCRs from class A, B, C, and F, Related to Figure 6. Buried polar residues in the transmembrane region are listed for six representative GPCR receptors in class specific residue numbering. Bolded residues indicate amino-acids with side-chains interacting with buried *de novo* predicted explicit waters through hydrogen bonds.

Class A: A2a receptor	Class B: CRF1 receptor	Class B: Glucagon receptor	Class C: mGlu1 receptor	Class C: mGlu5 receptor	Class F: SMO receptor
Buried Polar Residue	**Buried Polar Residue**	**Buried Polar Residue**	**Buried Polar Residue**	**Buried Polar Residue**	**Buried Polar Residue**
Y1.35	N1.43	Y1.43	S1.40	T1.54	H1.33
E1.39	**H1.47**	**Y1.47**	**S1.47**	S2.35	T1.43
N1.50	**S1.50**	S1.50	T1.54	Y2.50	T1.47
N2.40	R2.46	**N2.47**	S2.35	**T2.53**	**T1.50**
S2.45	N2.47	H2.50	Y2.50	Q3.32	T1.53
S2.50	H2.55	S2.56	T2.55	**R3.33**	N2.48
T3.36	**R2.60**	**K2.60**	S3.28	S3.39	**S2.55**
Q3.37	**N2.61**	D2.68	**Q3.32**	S3.43	W2.58
S3.38	W2.64	**Q3.37**	**R3.33**	**Y3.44**	Q2.61
S3.39	Q2.68	Y3.38	**S3.39**	S3.45	Y3.33
S3.42	T2.70	N3.43	**S3.40**	T3.49	Y3.34
D3.49	T3.33	**Y3.44**	**Y3.44**	K3.50	W3.42
W4.50	**Y3.36**	W3.46	S3.45	T3.51	**T3.47**
W4.53	**N3.37**	**E3.50**	T3.49	Q4.32	W3.50
N5.42	Y3.38	Y4.45	K3.50	Q4.43	**H3.51**
Y5.58	**H3.40**	W5.36	T3.51	N5.47	T4.49
S6.36	T3.42	R5.40	Q4.32	T5.55	T4.57
W6.48	**N3.43**	**N5.50**	Q4.43	E6.35	T5.68
H6.52	W3.46	R6.37	N5.47	T6.42	**T6.47**
N6.55	**E3.50**	S6.41	T5.55	**T6.46**	**H6.51**
Y7.36	W4.60	T6.42	E6.35	Y6.57	**E7.38**
S7.42	K4.64	**S7.46**	T6.42	S7.35	**N7.41**
H7.43	**D5.36**	S7.47	T6.46	**S7.37**	**T7.48**
N7.45	**Y5.39**	**Q7.49**	Y6.57	**S7.39**	**S7.53**
S7.46	**Q5.40**	Y7.57	T7.31	K7.51	W7.55
N7.49	N5.50		T7.32		**W7.57**
Y7.53	T6.42		**S7.37**		
	T6.52		**S7.39**		
	T6.53		K7.51		
	N7.42				
	E7.46				
	S7.47				
	Q7.49				
	S7.54				

Table S7. Data sets of protein-protein complex structures and mutations used for training and testing the hybrid energy function of SPaDES selected from SKEMPI, Related to STAR Methods text.

High-resolution hydrated set of protein-protein complex structures			
PDB	**Protein 1**	**Protein 2**	**Mutants considered for $\Delta\Delta G_{binding}$ predictions**
1ACB	Bovine alpha-chymotrypsin	Eglin C	I_L45D, I_L45G, I_L45S
1BRS	Barnase	Barstar	A_E73A, A_E73C, A_E73F, A_E73Q, A_E73S, A_E73W, A_E73Y, A_H102A, A_H102D, A_H102G, A_K27A, A_R83Q, A_R87A, D_D35A, D_D39A, D_T42A, D_W38F, D_W44F
1CSE	Subtilisin Carlsberg	Eglin C	I_L45E, I_L45G, I_L45I, I_L45S
1DQJ	HyHEL-63 Fab	HEW Lysozyme	A_N31A, A_N32A, A_S91A, A_Y50A, A_Y96A, B_D32A, B_W98A, B_Y33A, C_D101A, C_K96A, C_K97A, C_N93A, C_S100A, C_Y20A
1DVF	IgG1-kappa D1.3 Fv	E5.2 Fv	A_Y32A, B_E98A, B_R99A, B_W52A, B_Y101F, B_Y32A, D_D52A, D_H33A, D_Y98A
1JTG	TEM-1 beta-latamase	BLIP	A_E104A, A_E110A, A_M129A, A_P107A, A_Q99A, A_R243A, A_S130A, A_S235A, A_V103A, A_V216A, A_Y105A, B_D49A, B_F142A, B_F36A, B_H148A, B_H41A, B_K74A, B_S113A, B_S71A, B_W112A, B_W162A, B_Y143A, B_Y50A, B_Y53A
1TM1	Subtilisin BPN	Chymotrypsin inhibitor 2	I_E60A, I_E60S, I_M59A, I_M59G, I_M59K, I_R67A, I_R67C, I_T58A, I_T58D
1XD3	UCH-L3	Ubiquitin	B_H68N, B_I44A, B_K27A, B_K27R, B_L8A
2FTL	Bovine trypsin	BPTI	I_I18A, I_K15F, I_K15L, I_K15M, I_K15Q, I_K15S, I_K15W, I_K15Y
2G2U	SHV-1 beta-lactamase	BLIP	A_D104K, B_F142A, B_F36A, B_G141A, B_H148A, B_H41A, B_K74A, B_S113A, B_S39A, B_S71A, B_W112A, B_W150A, B_W162A, B_Y143A, B_Y50A, B_Y53A
2WPT	Colicin E2 immunity protein	Colicin E9 Dnase	A_D33A, A_S50A, B_F86A, B_N75A, B_S74A, B_S78A, B_S84A, B_T87A, B_V98A
3NPS	Membrane-type serine protease 1	S4 Fab	A_H143A, A_Q221A

Low- and high-resolution set of protein-protein complex structures			
PDB	**Protein 1**	**Protein 2**	**Mutants considered for $\Delta\Delta G_{binding}$ predictions**
1A22	Human growth hormone	hGH binding protein	A_C182A, A_E174A, A_F176A, A_F25A, A_H18A, A_H21A, A_I179A, A_K172A, A_L45A, A_N63A, A_P48A, A_P61A, A_R167A, A_R178A, A_S51A, A_S62A, A_T175A, A_Y164A, A_Y42A, B_C308A, B_C322A, B_D326A, B_D364A, B_I303A, B_I305A, B_I365A, B_K321A, B_N418A, B_P306A, B_Q416A, B_R243A, B_R270A, B_R271A, B_R417A, B_S302A, B_S324A, B_S419A, B_T301A, B_V371A, B_W276A, B_W304F
1A4Y	Ribonuclease inhibitor	Angiogenin	A_D435A, A_E344A, A_E401A, A_I459A, A_K320A, A_S289A, A_W261A, A_W263A, A_W318A, A_W375A, A_Y434A, A_Y434F, B_E108A, B_H114A, B_H13A, B_H84A, B_K40Q, B_Q12A, B_R31A, B_R5A
1ACB	Bovine alpha-chymotrypsin	Eglin C	I_L45D, I_L45G, I_L45S
1AK4	Cyclophilin A	HIV-1 capsid protein	D_A488G, D_A492G, D_A492V, D_G489A, D_P490A, D_V486A
1BRS	Barnase	Barstar	A_E73A, A_E73C, A_E73F, A_E73Q, A_E73S, A_E73W, A_E73Y, A_H102A, A_H102D, A_H102G, A_K27A, A_N58A, A_R59A, A_R59K, A_R83Q, A_R87A, D_D35A, D_D39A, D_T42A, D_W38F, D_W44F
1CBW	Bovine alpha-chymotrypsin	BPTI	I_G12A, I_G37A, I_I18A, I_K15A, I_K15E, I_K15F, I_K15G, I_K15H, I_K15M, I_K15N, I_K15Q, I_K15R, I_K15S, I_K15T, I_K15Y, I_P13A, I_R17A, I_T11A
1CSE	Subtilisin Carlsberg	Eglin C	I_L45E, I_L45G, I_L45I, I_L45S
1DAN	Factor VIIa	Tissue factor	H_M164A, T_D58A, T_D61A, T_F50A, T_F76A, T_G43A, T_I22A, T_I63A, T_K20A, T_K20R, T_K46A, T_K48A, T_L72A, T_N18A, T_Q37A, T_S16A, T_S47A, T_T17A, T_T21A, T_T60A, T_V64A, T_W14F, T_W45A, T_W45F, U_F140A, U_F147A, U_L133A, U_Q110A, U_T106A, U_T132A, U_V146A, U_V207A, U_W158F, U_Y156L, U_Y94A
1DQJ	HyHEL-63 Fab	HEW Lysozyme	A_N31A, A_N32A, A_S91A, A_Y50A, A_Y96A, B_D32A, B_W98A, B_Y33A, B_Y50A, B_Y53A, C_D101A, C_K96A, C_K97A, C_L75A, C_N93A, C_R21A, C_S100A, C_T89A, C_W63A, C_Y20A
1DVF	IgG1-kappa D1.3 Fv	E5.2 Fv	A_S93A, A_W92A, A_Y32A, A_Y49A, B_D58A, B_E98A, B_N56A, B_R99A, B_W52A, B_Y101F, B_Y32A, C_Y49A, D_D52A, D_H33A, D_I97A, D_Q100A, D_R100bA, D_Y98A
1EAW	Membrane-type serine protease 1	BPTI	A_F60eA, A_F94A, A_H143A, A_I41A, A_Q38A, A_Y60gA
1EMV	Colicin E9 immunity protein	Colicin E9 DNase	A_C23A, A_D51A, A_E41A, A_G49A, A_H46A, A_I53A, A_L33A, A_S48A, A_S50A, A_T27A, A_V34A, A_V37A, A_Y54A, B_F86A, B_K97A, B_N72A, B_N75A, B_R54A, B_S74A, B_S77A, B_S78A, B_S84A, B_T87A, B_V98A
1FCC	IgG1 MO61 Fc	B domain of Protein G	C_K28A, C_K31A, C_N35A, C_T25A, C_W43A
1FFW	Chemotaxis protein CheY	Chemotaxis protein CheA	A_Y106W, B_E171A, B_E178A, B_F214A, B_I216A
1GC1	HIV-1 gp120	CD4	C_E85A, C_H27A, C_K29A, C_K35A, C_K46A, C_L44A, C_N52A, C_Q25A, C_Q40A, C_R59A, C_S42A, C_T45A
1GCQ	Growth factor receptor-bound protein 2	VavS	C_A632G, C_P595A, C_P608A, C_P609A, C_P657A
1H9D	AML1 Runx1 Runt domain	Core-binding factor beta	B_L103A, B_N104A, B_Q67A, B_V4A
1IAR	Interleukin-4	Interleukin-4 receptor	A_E9Q, A_F82A, A_F82D, A_I5A, A_K12E, A_K12S, A_K84A, A_K84D, A_N89A, A_Q78A, A_Q78E, A_R85E, A_R88A, A_R88Q, A_T13A, A_T13D, A_T6A, A_T6D, A_W91A, A_W91D
1JCK	Beta-chain of 14.3.d	Staphylococcal enterotoxin C3	B_K103A, B_N60A, B_T20A, B_V91A, B_Y26A, B_Y90A
1JRH	mAbs A6	Interferon gamma receptor	H_H100A, H_W52A, H_W53A, H_Y58A, L_D28A, L_S93A, L_T94A, L_W92A, L_W96A, L_Y30A, L_Y91A, I_G50A, I_K47A, I_K47M, I_N48A, I_N53A, I_R84A, I_S54A, I_V51A, I_W56F, I_W56Y, I_W82A, I_W82F, I_W82Y, I_Y49A, I_Y49F
1JTG	IgG1-kappa D1.3 Fv	E5.2 Fv	A_E104A, A_E110A, A_M129A, A_P107A, A_Q99A, A_R243A, A_S130A, A_S235A, A_V103A, A_V216A, A_Y105A, B_D49A, B_F142A, B_F36A, B_H148A, B_H41A, B_K74A, B_S113A, B_S71A, B_W112A, B_W150A, B_W162A, B_Y143A, B_Y50A, B_Y53A
1KTZ	Transforming growth factor beta 3	TGF-beta type II receptor	A_V92I, B_E119A, B_E119Q, B_F30A, B_I50A, B_I53A, B_L27A, B_S52A, B_T51A, B_V77A
1MAH	Acetylcholinesterase	Fasciculin	A_D74N, A_F297I, A_F297Y, A_Y72N
1NMB	Subtype N9 neuraminidase	Antibody NC10	H_D56E, H_D56N, H_Y100aF, L_L94V, L_Y32F
1TM1	Subtilisin BPN	Chymotrypsin inhibitor 2	I_E60A, I_E60S, I_M59A, I_M59G, I_M59K, I_R65A, I_R67A, I_R67C, I_T58A, I_T58D, I_Y61A, I_Y61G
1VFB	IgG1-kappa D1.3 Fv	HEW Lysozyme	A_H30A, A_W92A, A_Y32A, A_Y49A, A_Y50A, B_D100A, B_W52A, B_Y32A, C_D18A, C_I124A, C_K116A, C_N19A, C_Q121A, C_S24A, C_T118A, C_V120A
1XD3	UCH-L3	Ubiquitin	B_H68N, B_I44A, B_K27A, B_K27R, B_L8A
1Z7X	Ribonuclease inhibitor	RNase A	W_D435A, W_E344A, W_I459A, W_K320A, W_S289A, W_W261A, W_W263A, W_W318A, W_W375A, W_Y434A, W_Y434F, W_Y437A
2FTL	Bovine trypsin	BPTI	I_I18A, I_K15F, I_K15L, I_K15M, I_K15Q, I_K15S, I_K15W, I_K15Y
2G2U	SHV-1 beta-lactamase	BLIP	A_D104K, B_F142A, B_F36A, B_G141A, B_H148A, B_H41A, B_K74A, B_S113A, B_S39A, B_S71A, B_W112A, B_W150A, B_W162A, B_Y143A, B_Y50A, B_Y51A, B_Y53A
2J0T	MMP1 Interstitial collagenase	Metalloproteinase inhibitor 1	D_C70S, D_S68A, D_S68E, D_T2A, D_T2R, D_T2S, D_V69I, D_V69T
2SIC	Subtilisin BPN	Streptomyces subtilisin inhibitor	I_M73A, I_M73G
2VLJ	HL-A2-flu	JM22	D_Q34A, D_S31A, D_S32A, E_D32A, E_D56A, E_I53L, E_I53V, E_Q58A, E_Q58E, E_S99A, E_Y101A
2WPT	Colicin E2 immunity protein	Colicin E9 Dnase	A_D33A, A_E41A, A_N34A, A_P56A, A_S50A, A_V37A, B_F86A, B_K97A, B_N72A, B_N75A, B_Q92A, B_R54A, B_S74A, B_S77A, B_S78A, B_S84A, B_T87A, B_V98A
3BK3	Bone morphogenetic protein-2	Crossveinless 2	C_I18A, C_I21A, C_I21R, C_I27A, C_I27R
3BN9	Membrane-type serine protease 1	E2 Fab	B_D217A, B_F94A, B_H143A, B_I41A, B_I60A, B_K224A, B_N95A, B_Q175A, B_T98A
3HFM	HyHEL-10	HEW Lysozyme	H_D32A, H_D32N, H_S31A, H_W98A, H_W98F, H_Y33A, H_Y33F, H_Y33L, H_Y50A, H_Y50L, H_Y53A, H_Y53F, H_Y53L, H_Y53W, H_Y58A, H_Y58F, H_Y58L, L_N31A, L_N31D, L_N31E, L_N32A, L_Q53A, L_Y50A, L_Y50L, L_Y96A, Y_D101A, Y_D101E, Y_D101G, Y_D101N, Y_D101Q, Y_D101R, Y_D101S, Y_H15A, Y_I98A, Y_K96A, Y_K97A, Y_K97E, Y_K97M, Y_L75A, Y_N93A, Y_R21A, Y_R21E, Y_R21G, Y_R21H, Y_R21K, Y_R21M, Y_R21N, Y_R21Q, Y_R21W, Y_S100A, Y_W63A, Y_Y20A, Y_Y20F, Y_Y20L
3NPS	Membrane-type serine protease 1	S4 Fab	A_F97A, A_H143A, A_I41A, A_Q175A, A_Q221A, A_Y146A
4CPA	Carboxypeptidase A	Potato carboxypeptidase inhibitor	I_P36G, I_V38A, I_V38G, I_V38L, I_Y37F, I_Y37G