

## **Supplementary data for: “Prediction of the effect of point mutations on Drug-Protein binding affinity”**

---

<b>Table S1. Mapping of PDB to ECIF atom types for protein structures.</b>	<b>2</b>
<b>Table S2. RDKit ligand descriptors</b>	<b>3</b>
<b>Table S3. RDKit Mutation descriptors</b>	<b>3</b>
<b>Table S4. Full list of PDB IDs of structures forming the Dataset.</b>	<b>3</b>
<b>Table S5. Hyperparameters considered for each Machine learning Model</b>	<b>4</b>
<b>Table S6. PDB ID distribution over Train, Test and Validation sets.</b>	<b>4</b>

PDB Residue	PDB Atom	ECIF Atom	PDB Residue	PDB Atom	ECIF Atom	PDB Residue	PDB Atom	ECIF Atom	PDB Residue	PDB Atom	ECIF Atom	PDB Residue	PDB Atom	ECIF Atom
ALA	C	C;4;3;0;0;0	CYS	C	C;4;3;0;0;0	HIS	C	C;4;3;0;0;0	MET	C	C;4;3;0;0;0	THR	C	C;4;3;0;0;0
	CA	C;4;3;1;0;0		CA	C;4;3;1;0;0		CA	C;4;3;1;0;0		CA	C;4;3;1;0;0			
	CB	C;4;1;3;0;0		CB	C;4;2;2;0;0		CB	C;4;2;2;0;0		CB	C;4;3;1;0;0			
	N	N;3;2;1;0;0		N	N;3;2;1;0;0		CD2	C;4;2;1;1;1		CE	C;4;1;3;0;0			
	O	O;2;1;0;0;0		O	O;2;1;0;0;0		CE1	C;4;2;1;1;1		CG	C;4;2;2;0;0			
OXT	O;2;1;0;0;0	OXT	O;2;1;0;0;0	CG	C;4;3;0;1;1	N	N;3;2;1;0;0	O	O;2;1;0;0;0					
C	C;4;3;0;0;0	SG	S;2;1;1;0;0	N	N;3;2;1;0;0	O	O;2;1;0;0;0	OXT	O;2;1;0;0;0					
CA	C;4;3;1;0;0	CA	C;4;3;1;0;0	ND1	N;3;2;0;1;1	OXT	O;2;1;0;0;0	OXT	O;2;1;0;0;0					
CB	C;4;2;2;0;0	CB	C;4;2;2;0;0	NE2	N;3;2;1;1;1	SD	S;2;2;0;0;0	C	C;4;3;0;0;0					
CD	C;4;2;2;0;0	CD	C;4;3;0;0;0	O	O;2;1;0;0;0	C	C;4;3;0;0;0	CA	C;4;3;1;0;0					
CG	C;4;2;2;0;0	CG	C;4;2;2;0;0	OXT	O;2;1;0;0;0	CA	C;4;3;1;0;0	CB	C;4;2;2;0;0					
CZ	C;6;3;0;0;0	GLN	CG	C;4;2;2;0;0	C	C;4;3;0;0;0	CB	C;4;2;2;0;0	CD1	C;4;2;1;1;1				
N	N;3;2;1;0;0	N	N;3;2;1;0;0	CA	C;4;3;1;0;0	CD1	C;4;2;1;1;1	CD2	C;4;3;0;1;1					
NE	N;4;2;1;0;0	NE2	N;3;1;2;0;0	CB	C;4;3;1;0;0	CD2	C;4;2;1;1;1	CE2	C;4;3;0;1;1					
NH1	N;4;1;2;0;0	O	O;2;1;0;0;0	CD1	C;4;1;3;0;0	CE1	C;4;2;1;1;1	CE3	C;4;2;1;1;1					
NH2	N;4;1;2;0;0	OE1	O;2;1;0;0;0	CG1	C;4;2;2;0;0	CE2	C;4;2;1;1;1	CG	C;4;3;0;1;1					
O	O;2;1;0;0;0	OXT	O;2;1;0;0;0	CG2	C;4;1;3;0;0	CG	C;4;3;0;1;1	CH2	C;4;2;1;1;1					
OXT	O;2;1;0;0;0	C	C;4;3;0;0;0	N	N;3;2;1;0;0	CZ	C;4;2;1;1;1	CZ2	C;4;2;1;1;1					
C	C;4;3;0;0;0	CA	C;4;3;1;0;0	O	O;2;1;0;0;0	N	N;3;2;1;0;0	CZ3	C;4;2;1;1;1					
CA	C;4;3;1;0;0	CB	C;4;2;2;0;0	OXT	O;2;1;0;0;0	O	O;2;1;0;0;0	N	N;3;2;1;0;0					
CB	C;4;2;2;0;0	CD	C;5;3;0;0;0	CA	C;4;3;1;0;0	OXT	O;2;1;0;0;0	NE1	N;3;2;1;1;1					
CG	C;4;3;0;0;0	CG	C;4;2;2;0;0	CB	C;4;2;2;0;0	C	C;4;3;0;0;0	O	O;2;1;0;0;0					
N	N;3;2;1;0;0	N	N;3;2;1;0;0	CD1	C;4;1;3;0;0	CA	C;4;3;1;0;1	OXT	O;2;1;0;0;0					
ND2	N;3;1;2;0;0	O	O;2;1;0;0;0	CD2	C;4;1;3;0;0	CB	C;4;2;2;0;1	C	C;4;3;0;0;0					
O	O;2;1;0;0;0	OE1	O;2;1;0;0;0	CG	C;4;3;1;0;0	CD	C;4;2;2;0;1	CA	C;4;3;1;0;0					
OD1	O;2;1;0;0;0	OE2	O;2;1;0;0;0	N	N;3;2;1;0;0	CG	C;4;2;2;0;1	CB	C;4;2;2;0;0					
OXT	O;2;1;0;0;0	OXT	O;2;1;0;0;0	O	O;2;1;0;0;0	N	N;3;3;0;0;1	CD1	C;4;2;1;1;1					
C	C;4;3;0;0;0	C	C;4;3;0;0;0	O	O;2;1;0;0;0	O	O;2;1;0;0;0	CD2	C;4;2;1;1;1					
CA	C;4;3;1;0;0	CA	C;4;2;2;0;0	OXT	O;2;1;0;0;0	OXT	O;2;1;0;0;0	CE1	C;4;2;1;1;1					
CB	C;4;2;2;0;0	N	N;3;2;1;0;0	C	C;4;3;0;0;0	CA	C;4;3;1;0;0	CE2	C;4;2;1;1;1					
CG	C;5;3;0;0;0	O	O;2;1;0;0;0	CA	C;4;3;1;0;0	CB	C;4;2;2;0;0	CG	C;4;3;0;1;1					
N	N;3;2;1;0;0	OXT	O;2;1;0;0;0	CD	C;4;2;2;0;0	N	N;3;2;1;0;0	CZ	C;4;3;0;1;1					
O	O;2;1;0;0;0	CE	C;4;2;2;0;0	CE	C;4;2;2;0;0	O	O;2;1;0;0;0	N	N;3;2;1;0;0					
OD1	O;2;1;0;0;0	CG	C;4;2;2;0;0	CG	C;4;2;2;0;0	OG	O;2;1;1;0;0	O	O;2;1;0;0;0					
OD2	O;2;1;0;0;0	N	N;3;2;1;0;0	N	N;3;2;1;0;0	OXT	O;2;1;0;0;0	OH	O;2;1;1;0;0					
OXT	O;2;1;0;0;0	NZ	N;4;1;3;0;0	O	O;2;1;0;0;0	OXT	O;2;1;0;0;0	OXT	O;2;1;0;0;0					
		O	O;2;1;0;0;0	OXT	O;2;1;0;0;0			C	C;4;3;0;0;0					
				</										

**Table S2. RDKit ligand descriptors**


---

BalabanJ, BertzCT, Chi0, Chi0n, Chi0v, Chi1, Chi1n, Chi1v, Chi2n, Chi2v, Chi3n, Chi3v, Chi4n, Chi4v, EState\_VSA1, EState\_VSA10, EState\_VSA11, EState\_VSA2, EState\_VSA3, EState\_VSA4, EState\_VSA5, EState\_VSA6, EState\_VSA7, EState\_VSA8, EState\_VSA9, ExactMolWt, FpDensityMorgan1, FpDensityMorgan2, FpDensityMorgan3, FractionCSP3, HallKierAlpha, HeavyAtomCount, HeavyAtomMolWt, Kappa1, Kappa2, Kappa3, LabuteASA, MaxAbsEStateIndex, MaxEStateIndex, MinAbsEStateIndex, MinEStateIndex, MolLogP, MolMR, MolWt, NHOHCount, NOCount, NumAliphaticCarbocycles, NumAliphaticHeterocycles, NumAliphaticRings, NumAromaticCarbocycles, NumAromaticHeterocycles, NumAromaticRings, NumHAcceptors, NumHDonors, NumHeteroatoms, NumRotatableBonds, NumSaturatedCarbocycles, NumSaturatedHeterocycles, NumSaturatedRings, NumValenceElectrons, PEOE\_VSA14, RingCount, SMR\_VSA1, SMR\_VSA10, SMR\_VSA2, SMR\_VSA3, SMR\_VSA4, SMR\_VSA5, SMR\_VSA6, SMR\_VSA7, SMR\_VSA9, SlogP\_VSA1, SlogP\_VSA10, SlogP\_VSA11, SlogP\_VSA12, SlogP\_VSA2, SlogP\_VSA3, SlogP\_VSA4, SlogP\_VSA5, SlogP\_VSA6, SlogP\_VSA7, SlogP\_VSA8, TPSA, VSA\_EState1, VSA\_EState10, VSA\_EState2, VSA\_EState3, VSA\_EState4, VSA\_EState5, VSA\_EState6, VSA\_EState7, VSA\_EState8, VSA\_EState9, fr\_Al\_COO, fr\_Al\_OH, fr\_Al\_OH\_noTert, fr\_ArN, fr\_Ar\_N, fr\_Ar\_NH, fr\_Ar\_OH, fr\_COO, fr\_COO2, fr\_C\_O, fr\_C\_O\_noCOO, fr\_C\_S, fr\_HOCCN, fr\_Imine, fr\_NH0, fr\_NH1, fr\_NH2, fr\_N\_O, fr\_Ndealkylation1, fr\_Ndealkylation2, fr\_Nhprrrole, fr\_SH, fr\_aldehyde, fr\_alkyl\_carbamate, fr\_alkyl\_halide, fr\_allylic\_oxid, fr\_amide, fr\_amidine, fr\_aniline, fr\_aryl\_methyl, fr\_azo, fr\_barbitur, fr\_benzene, fr\_bicyclic, fr\_dihydropyridine, fr\_epoxide, fr\_ester, fr\_ether, fr\_furan, fr\_guanido, fr\_halogen, fr\_hdrzine, fr\_hdrzone, fr\_imidazole, fr\_imide, fr\_isocyan, fr\_isothiocyan, fr\_ketone, fr\_ketone\_Topliss, fr\_lactam, fr\_lactone, fr\_methoxy, fr\_morpholine, fr\_nitrile, fr\_nitro, fr\_nitro\_arom, fr\_nitroso, fr\_oxazole, fr\_oxime, fr\_para\_hydroxylation, fr\_phenol, fr\_phenol\_noOrthoHbond, fr\_piperdine, fr\_piperzine, fr\_priamide, fr\_pyridine, fr\_quatN, fr\_sulfide, fr\_sulfonamd, fr\_sulfone, fr\_term\_acetylene, fr\_tetrazole, fr\_thiazole, fr\_thiocyan, fr\_thiophene, fr\_urea, qed

---

**Table S3. RDKit Mutation descriptors**


---

Hydropathy\_change, Volume\_change, Polarity\_change, PK\_side\_chain\_change, Prct\_exposed\_residues\_change, Hydrophilicity\_change, Accessible\_surface\_area\_change, Local\_flexibility\_change, Accessible\_surface\_area\_folded\_change, Refractivity\_change, Mass\_change, Solvent\_exposed\_area\_change, Mutation\_distance\_ligand

---

**Table S4. Full list of PDB IDs of structures forming the Dataset.**


---

1A4H, 1AKR, 1AKT, 1AKU, 1AKV, 1AKW, 1AMK, 1AMW, 1B74, 1BF3, 1C7E, 1CNQ, 1DF8, 1DLR, 1E2K, 1EGH, 1FLM, 1FLV, 1FSG, 1G6S, 1G6T, 1GZ4, 1GZF, 1H0A, 1HK1, 1HNN, 1IYK, 1K3L, 1KC7, 1KDK, 1L7E, 1LO7, 1LO8, 1N7I, 1N7J, 1NJB, 1NJE, 1O0N, 1OGZ, 1OH0, 1OPJ, 1OSS, 1OWB, 1PIN, 1PIO, 1P7T, 1PWM, 1QY1, 1SYH, 1THY, 1TPZ, 1TSL, 1TSM, 1U72, 1US0, 1W4Q, 1W96, 1WCQ, 1YYZ, 1YZ3, 1Z95, 2AMA, 2C80, 2CA8, 2CC7, 2CCB, 2CCC, 2DOR, 2E7F, 2FZD, 2G70, 2G72, 2GMK, 2H4N, 2HA3, 2IKI, 2INE, 2INZ, 2ITY, 2J27, 2JBZ, 2JDY, 2JGS, 2OBF, 2ONY, 2P4T, 2PDG, 2PQL, 2PZN, 2Q2A, 2Q88, 2Q89, 2QEH, 2QFQ, 2QFS, 2QFT, 2QFU, 2RDE, 2TDM, 2V2C, 2V5V, 2VQZ, 2VWC, 2W3A, 2ZGA, 3AH4, 3AJ5, 3AM3, 3AQT, 3C3U, 3CKZ, 3CL0, 3CS9, 3DGL, 3DGN, 3DT4, 3DTB, 3E2Q, 3E2R, 3E2S, 3E6K, 3E6V, 3E9B, 3EEB, 3EIG, 3FQL, 3FRE, 3FS6, 3G0E, 3H2K, 3HU1, 3HU2, 3HU3, 3HY8, 3I73, 3I8A, 3IJW, 3MI3, 3NOH, 3NTY, 3OKC, 3OXZ, 3QFX, 3QGT, 3RG9, 3RY2, 3S3V, 3UE4, 3UJ9, 3UJB, 3UM8, 3ZY2, 4B7J, 4B7N, 4WA9, 4XEY

---

**Table S5. Hyperparameters considered for each Machine learning Model**

Model	Tested Hyperparameters and possible values
Random Forest Classifier	Maximum depth (“max_depth”): 2, 8, 16, 25, 35, 50, 70
	Number of decision tree regressors to be built (“n_estimators”): 32, 64, 128, 256, 420, 600, 650
	Fraction of features to look at for the best split (“max_features”): 1, 5, 10, 15, 20, 35, 50
	Minimum number of samples required to split an internal node (“min_samples_split”): 1, 2, 3, 5, 8, 12, 15
Gradient Boosting Regressor	Maximum depth (“max_depth”): 2, 8, 16, 25, 35, 50, 70
	Number of decision tree regressors to be built (“n_estimators”): 32, 64, 128, 256, 420, 600
	Minimum number of samples required to split an internal node (“min_samples_split”): 1, 2, 3, 5, 8, 12, 15
	Speed at which the model learns (“learning_rate”) learning_rate = 0.001, 0.005, 0.02, 0.05
	Loss function to be optimized (“loss”): squared_error, absolute_error, huber, quantile
*All remaining hyperparameters were set as default according to the Scikit-learn Python library (0.22.1)	

**Table S6. PDB ID distribution over Train, Test and Validation sets.**

Training Set	Testing Set	Validation Set
1A4H, 1AKV, 1AMW, 1B74, 1CNQ, 1DF8, 1DLR, 1EGH, 1FLV, 1FSG, 1G6S, 1G6T, 1GZ4, 1GZF, 1HK1, 1HNN, 1IYK, 1K3L, 1KDK, 1L7E, 1LO7, 1LO8, 1N7I, 1N7J, 1NJB, 1NJE, 1O0N, 1OGZ, 1OH0, 1OSS, 1PWM, 1SYH, 1THY, 1TSL, 1TSM, 1U72, 1US0, 1W4Q, 1W96, 1YZ3, 2AMA, 2C80, 2CA8, 2CC7, 2CCB, 2CCC, 2DOR, 2FZD, 2G70, 2G72, 2GMK, 2HA3, 2IKI, 2INE, 2INZ, 2JBZ, 2ONY, 2PDG, 2PZN, 2Q2A, 2Q88, 2Q89, 2QFS, 2QFT, 2RDE, 2TDM, 2V5V, 2VQZ, 2VWC, 2W3A, 2ZGA, 3AH4, 3AJ5, 3AQT, 3C3U, 3DGL, 3DGN, 3DT4, 3DTB, 3E2Q, 3E2R, 3EEB, 3EIG, 3FRE, 3FS6, 3H2K, 3I73, 3I8A, 3IJW, 3MI3, 3N0H, 3NTY, 3QFX, 3QGT, 3RG9, 3RY2, 3S3V, 3UJ9, 3UJB, 3UM8, 4B7J, 4B7N	1AMK, 1BF3, 1E2K, 1FLM, 1H0A, 1KC7, 1P7T, 1QY1, 1TPZ, 1WCQ, 2E7F, 2H4N, 2ITY, 2PQL, 2QEH, 3AM3, 3FQL, 3G0E, 3HY8, 3OKC, 3ZY2	1OPJ, 3CS9, 3OXZ, 3UE4, 4WA9, 4XEY