Supporting information

Computation-guided engineering of distal mutations in an artificial enzyme

Fabrizio Casilli [1], Miquel Canyelles-Niño [2], Gerard Roelfes [1]* and Lur Alonso-Cotchico [2]*

[1] Stratingh Institute for Chemistry, University of Groningen, 9747 AG, Groningen, The Netherlands
[2] Zymvol Biomodeling S.L., Barcelona, Spain

Contents

General considerations

Chemicals were purchased from *Sigma-Aldrich* (UK) and *TCI* (Belgium/Japan), and used without further purification unless specified. The unnatural amino acid p-azido phenylalanine (pAzF) was purchased as the enantiopure freebase from *Iris Biotech* (DE). Plasmid pEVOL-pAzFRS.2.t1 was a gift from Farren Isaacs (Addgene plasmid # 73546).[1] *E. coli* strains NEB10β and BL21(DE3) (*New England Biolabs*) were used for cloning and expression. Unless otherwise noted, cells were made chemically competent using the Inoue protocol.[2] DNA primers were synthesised by *Eurofins genomics* (Germany) and *Sigma-Aldrich* (UK). Plasmid Purification (QIAprep Spin Miniprep Kit) and PCR clean-up kits (QIAquick PCR Purification Kit) were obtained from *QIAGEN* (Germany). Sanger DNA sequencing was carried out by *Eurofins genomics* (Germany). *Phusion* polymerase, HF buffer 10X, deoxynucleotide (dNTP) solution mix, dimethyl sulfoxide (DMSO), *DpnI* and 2X HiFi Assembly mastermix were purchased from *New England Biolabs*. *Pfu Turbo* DNA polymerase was purchased from *Agilent* (USA). Strep-tactin columns (Strep-Tactin® Superflow® high capacity) and Desthiobiotin were purchased from *IBA-Lifesciences* (Germany). Concentrations of DNA and protein solutions were determined based on the absorption at 260 nm or 280 nm on a *Thermo Scientific* Nanodrop 2000 UV-Vis spectrophotometer. Molar extinction coefficients of protein mutants were approximated using the ProtParam Expasy web server https://web.expasy.org/protparam/. UV/Vis absorption spectra and kinetic assays were recorded at 25 °C on a Jasco V-660 spectrophotometer unless specified. CD spectra were recorded on a Jasco J-810 CD Spectrometer using 1 mm quartz cuvettes. High-Resolution Mass Spectrometry (HR-MS) measurements were performed using a *Thermo* LTQ Orbitrap XL. Low-Resolution Mass Spectrometry (LR-MS) measurements were performed using a *Waters* Acquity H-class UPLC with *Waters* Xevo G2 QTOF. Theoretical molecular weights of proteins were calculated using the Thermo FreeStyle™ application.

Methods

General PCR protocol for site-directed mutagenesis

PCR was performed in 50 µL total volume containing HF buffer (5x stock solution), dNTPs mixture (0.2 mM each dNTP), DNA template (10-30 ng), primers (0.3 µM), DMSO (3 % v/v) and *Phusion* polymerase (1 U). The mutations were introduced using the appropriate primer pairs (see table Primer sequence). The reactions were then briefly centrifuged, transferred to a thermocycler (*Eppendorf Mastercycler*) and subjected to the following Touch-Down PCR program: (1) initial denaturation (95 °C, 6 min), (2) 20 cycles of denaturation (95 °C, 30 s), annealing (from 65 °C to 45 °C with steps of -1 °C/cycle, 30 s), and extension (72 °C, 2 min and 10 s), (3) 10 cycles of denaturation (95 °C, 30 s), annealing (45 °C, 30 s), and extension (72 °C, 2 min and 10 s), (4) a final extension (72 °C, 2 min). The resulting PCR product was digested by addition of 20 U of *DpnI* and incubated for 1 h at 37 °C. Afterwards, 5 µL of PCR reaction was transformed via heat-shock at 42 °C for 45 s in chemically competent *E. coli* NEB10β cells for storage or in chemically competent *E. coli* BL21(DE3) containing the plasmid pEVOL-pAFRS2.t1 for protein expression. The transformation mixture was spread onto Luria-Bertani (LB)-Agar plates containing ampicillin (100 µg/mL) for *E. coli* NEB10β cells or ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL) for *E. coli* BL21(DE3) cells. After incubation overnight at 37 °C, single colonies were picked and transferred into 5 mL of LB supplemented with the appropriate antibiotic. After incubation overnight at 37 °C, shaking at 135 rpm, the densely grown culture was used to prepare a glycerol stock (17% v/v glycerol) and for miniprep (kit from *QIAGEN* (Germany)) and Sanger

sequencing of the mutated plasmid. The site-directed mutagenesis for the preparation of the library of distal mutants followed the same Touch-Down protocol with minor modifications: the extension time was doubled to account for the speed of PfuTurbo DNA polymerase. The consensus sequence was prepared in two consecutive steps: (1) Gibson assembly[3] to incorporate 5 point mutations (A11G, A48G,T52P, F54L, M89L), (2) Quikchange site-directed mutation to incorporate one mutation (V28G). Gibson assembly was performed on three fragments called BB, F1 and F2 (3803 bp, 161 bp, and 127 bp respectively). Fragment BB was obtained with standard PCR protocol using the primers *Cons_M89L_fwd* and *Cons_A11G_rv*. Fragment F1 was amplified with primers *Cons_A11G_fwd* and Cons_A48G_T52P_F54L_rv. Fragment F2 was amplified with primers *Cons_A48G_T52P_F54L_fwd* and *Cons_M89L_rv*. Gibson assembly was performed with a molar ratio 1:3:3 of BB:F1:F2 with 75 ng of BB in 20 µL (10 µL 2X HiFi Assembly mastermix and 10 µL of DNA fragments mixture) at 50 °C for 1 h. The assembly product (5 µL) was transformed into *E. coli* NEB10β and the plasmid DNA was harvested as described before. Upon confirmation of the mutations via Sanger sequencing, the plasmid obtained was used as template for the next step of mutagenesis to incorporate mutation V28G. Standard PCR protocol was employed using *Cons_V28G_fwd* and *Cons_V28G_rv* as primers. Mutation was confirmed by Sanger sequencing.

Library preparation in 96-well plate

Transformants, either from agar plate or glycerol stock, were inoculated in a 96-deep well plate in 1 mL of LB supplemented with ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL). Every plate contained the following controls, each in triplicate: LmrR_RMH, LmrR_V15pAF and LmrR wild-type. The resulting plate was incubated overnight at 37 °C shaking at 900 rpm (*Titramax 1000 & Incubator 1000, Heidolph*). Afterwards, 50 µL of the overnight culture was used to inoculate at least 3 plates containing 1100 µL of fresh LB supplemented with antibiotics and incubated for 4 h at 37 °C shaking at 900 rpm. Subsequently, protein production was induced by addition of 50 µL LB media, containing IPTG (1 µL of a 1 M stock solution), arabinose (1 µL of a 20% arabinose stock solution) and p-azidophenylalanine (pAzF) at a concentration of 20 mM (final concentrations: IPTG = 0.8 mM, arabinose = 0.017%, pAzF = 1 mM). To avoid precipitation of the unnatural amino acid, pAzF was dissolved by addition of 1 equivalent of base (1 M NaOH) prior to addition to the LB media. Plates were then incubated at 30 °C for 16 hours while shaking (900 rpm) and harvested by centrifugation (3000 x g at 4 °C for 20 minutes). The pellet was washed in 500 µL of Buffer R (50 mM Na_2HPO_4, 150 mM NaCl, pH =7.5) and centrifuged again (3000 x g at 4 °C for 20 minutes) to remove any residues of LB from cells and froze overnight at -20 °C. Lysis of the cells was performed by resuspension in 300 µL of lysis buffer containing Buffer R, protease inhibitor cocktail (*Roche cOmplete*), lysozyme (1 mg/mL), DNase I (0.1 mg/mL) and $MgCl_2$ (10 mM). Cells were incubated for 2 h at 30 °C at 900 rpm to ensure full cell lysis was achieved. Afterwards, the ncAA pAzF was reduced to *p*-aminophenyl alanine (pAF) via Staudinger reduction by supplying the lysate with 30 µL of TCEP (tris(2-carboxyethyl)phosphine) stock solution (100 mM in Buffer R, 10 mM final concentration). Reduction was performed at room temperature for at least 1 h shaking at 900 rpm. Cell debris were removed with centrifugation for 45 min at 4 °C, 3000 x g. The cleared lysate was immediately assayed for activity.

Combinatorial library preparation and screening

The combinatorial library was made following the multichange isothermal (MISO) assembly procedure.[4] The PCR was performed following the general protocol described with plasmid

pET17b_LmrR_RMH_F54L as template (12 ng). For the backbone fragment, the primers used were *I62NDT_rv* and *N88VHG_fwd* (0.3 µM, theoretically corresponding to 25 nM for each codon) supplemented with the primers that would include the I62W mutation and ensure the presence of N88Q in the diversity, namely *I62W_rv_2* and *N88Q_fwd_2* (25 nM each). For the insert fragment, primers used were *N88VHG_rv* and *I62NDT_fwd* (0.3 µM) supplemented with primers *N88Q_rv_2* and *I62W_fwd_2* (25 nM each). For the backbone fragment, amplification was achieved according to the following program: (1) initial denaturation (98 °C, 3 min), (2) 20 cycles of denaturation (98 °C, 30 s), annealing (50 °C, 30 s) and extension (72 °C 2 min), (3) final extension (72 °C, 10 min). The insert fragment was amplified as follows: (1) initial denaturation (98 °C, 3 min), (2) 20 cycles of denaturation (98 °C, 30 s), annealing (58 °C, 30 s) and extension (72 °C 20 s), (3) final extension (72 °C, 10 min). Both the backbone and insert fragment PCR products (5 µL) were run on 1% agarose gel to confirm amplification (data not shown, backbone was 3965 bp, insert was 98 bp). PCR clean-up of the remaining 45 µL was performed. Gibson assembly of backbone and fragment was performed using a molar ratio of 1:5 respectively (80:10 ng) in 20 µL (10 µL 2X HiFi Assembly mastermix and 10 µL of DNA fragments mixture) at 50 °C for 1 h. The assembly mix (5 µL) was electroporated in *E. coli* Bl21(DE3) cells (made electrocompetent following the protocol in [ref]) already harbouring the plasmid pEVOL-pAFRS2.t1. Of the transformants, 440 single cells were picked (corresponding to 3.8-times the theoretical diversity in the library, thereby ensuring >95% coverage) and inoculated in 1 mL LB supplemented with antibiotics in 96-deep well plate. Then, the standard library preparation and 96-well plate hydrazone formation activity assay protocols were followed. The mutants showing more than 1.2-fold improvement from the parent were picked for sequencing and characterization in purified format.

Computational methods

Structure preparation. The crystallographic X-ray data for LmrR_RMH (unpublished structure) was used as the starting point for all the computational simulations. Due to their high flexibility, the structure of the loops formed by residues 70-74 in chain A and 70-73 in chain B was not solved; thus, they were grafted by loop reconstruction with Yasara.[5] Mutations F54L and N88Q were introduced using the AmberTools' pdb4amber tool.[6]

Analysis of protein trajectories from MD simulations. Trajectory analysis was carried out with Python 3.11.7 and a combination of MDTraj v1.9.9,[8] PyTraj v2.0.6,[9,10] MDAnalysis v2.7.0,[11,12] NumPy v1.26.3,[13] RCBS.py suite v1.2.0,[14] Pandas v2.2.0.[15] Plots were drawn with Matplotlib v3.8.2[16] and Seaborn v0.13.2.[17] Visual inspection of the structures was performed with UCSF ChimeraX (v 1.7)[18] and Visual Molecular Dynamics (VMD) v1.9.3 software. Alignment of the structures and calculation of RMSD for static structures was performed with the Matchmaker tool embedded in ChimeraX.

The root-mean-square deviation (RMSD) was measured along the simulations using as a reference the first frame for each replicate, while the root-mean-square fluctuation (RMSF) was first calculated using as a reference the average structure of each replicate and, consequently, averaged among all replicates to obtain a unique profile for each mutant.

Cluster analysis was performed on the full trajectory of the parent and both the mutants without solvent using the k-means algorithm. To identify the correct number of clusters describing the system, the quality metrics Davies-Bouldin index (DBI), pseudo-F-statistic (pSF), and the ratio between the sum of squares regression and the total sum of squares (SSR/SST) were analysed.[19] The suitable number of

clusters was selected based on the lowest DBI value, highest pSF, and stable SSR/SST. Analysis of clusters in the hydrophobic pocket was obtained by selecting the catalytic non-canonical amino acid pAF together with the central structural residue W96 for both monomers. Clustering of the backbone included the alpha carbon of the whole protein, excluding the more flexible loops (i.e. G1-P5, G68-R76, A108-K116, G1'-P5', G68'-R76', K110'-K116'). Clustering of the DNA-binding site (DBS) included the residues in helices α2 and α3 (i.e. from Y27 to I63); the region of the beta-wing was purposefully left out of the selection due to its intrinsic elevated dynamics, which would have affected the formation of structurally consistent clusters. In the case of the hydrophobic pocket clustering of LmrR_RMH, following pSF, DBI, and SSR/SST trends would lead to the selection of five clusters as the optimum amount; however, a too small within-cluster standard deviation suggested that fewer clusters would better compare to the metrics extracted for the mutants. In this case, it was adjusted to 3. Clustering was also performed based on the orientation of pAF15/pAF15' and W96/W96' (pAF+W96). Finally, all those residues belonging to the strongest allosteric pathway that connected F54L to pAF in both monomers (i.e. residues pAF15, L17, T13, N14, F54(L), F54(L)', I53', R10', N12', T13', N14', pAF15') were selected for cluster analysis.

The three states were found by visually inspecting representative structures of the hydrophobic pocket clusters. The allocation of a frame to each of the three states was obtained by monitoring the orientation of R92 along the trajectory. Each frame was inspected for four geometric distances:

(i) between nitrogen NH1 of R92 guanidinium side chain and the carbonyl oxygen of N14 (distA),

(ii) between nitrogen NH1 of R92' guanidinium side chain and the carbonyl oxygen of N14' (distB),

(iii) between nitrogen NH1 of R92 guanidinium side chain with the carboxylic oxygen of D100' (distC),

(iv) between nitrogen NH1 of R92' guanidinium side chain with the carboxylic oxygen of D100 (distD).

More specifically, the structure of LmrR in each frame of the simulation was classified as having:

(i) a *trans* conformation if distA was lower than 8 Å, distB was higher than 10 Å, distC was higher than 11 Å and distD was lower than 8 Å.

(ii) a *cis-back* conformation if distA was lower than 8 Å, distB lower than 8 Å, distC higher than 11 Å, and distD higher than 11 Å.

(iii) a *cis-front* conformation if distA was higher than 8 Å, distB was higher than 10 Å, distC was lower than 11 Å and distD was lower than 8 Å.

Additionally, the formation of hydrogen bonds was measured along the trajectory using the HydrogenBondAnalysis class of MDAnalysis with default parameters. The radius of gyration was used to monitor the level of compactness of LmrR and was measured on the whole protein with the exception of the termini and highly flexible loops (residues G1-E3, D60-T82, K110-K116, G1'-E3', D60'-T82', K110'-K116').

The contact frequency analysis was performed using EMDA v1.0.0a4[20]. Two residues were considered in contact when they were found closer than 3 Å. Contacts were registered every 100 ps of simulation for each replicate when handling the full trajectories, or every 10 ps (i.e. every frame) when the contacts were analysed for each of the identified states. The measured contact frequency for each

mutant was then compared to the one of the parent, using a threshold of 30% of relative frequency to capture this contact as relevant.

Dynamical network analysis. In order to identify important long-range interactions and understand the allosteric networks in LmrR, the dynamical network analysis[21] was employed. This approach uses a generalised correlation of motion and short path betweenness centrality to define and rank the edges among the residues. The dynamic network analysis was performed using the workflow established by Melo and co-workers[21] (dynetan v2.2.0), where the Floyd–Warshall algorithm was used to identify the optimum paths. An interaction was considered acceptable if the distance between two nodes was below 4.5 Å for at least 75% of the trajectory. The number of windows was set to 3 to capture small conformational changes. Those interactions that were more commonly found among all the windows were considered.

Supporting figures and tables

Prediction of long-range mutations

Table S1: List of all the positions predicted to be relevant hotspots in the conformational dynamics of LmrR, together with the mutations experimentally analysed. Those positions that were predicted but not tested experimentally are highlighted in red. The mutations belonging to the consensus are underlined. The mutations belonging to the consensus that were not tested as single point mutations, but were only used in combination with others, are highlighted in yellow.

Original residue	**Amino acid mutations proposed**		
E7	D		
L9	R	K	
R10	Q	K	A
A11	G		
Q12	V	E	
T13	L	I	
N14	E		
I16	M	L	C
N19			
V20	L	I	
K22	S	E	
Q23	L		
G24	E		
D25	P	E	
N26	S	R	
V28	G		
G30	S		
I32	V	T	
Q34	R	E	
A38	L		
S39	G		
E42	V		
L45	M	I	
N46			
E47	P		
A48	G		
T52	P		

I53	V	S	L
F54	L		
K55	S	R	
E58	Q		
I62	Y	W	L
S64	T	E	
S65	T	G	
Y66	R		
R75	P		
R80	S		
T82	S		
E83			
H86	R	K	E
E87			

N88	Q	E	
M89	L		
R90	Q	A	
A92			
E94	Q	A	
S95	E	D	
S97	Q	K	
R98			
V99	L		
K101	R		
I102	V		
I103			
E107	D		
K110	G		

Multi sequence alignment

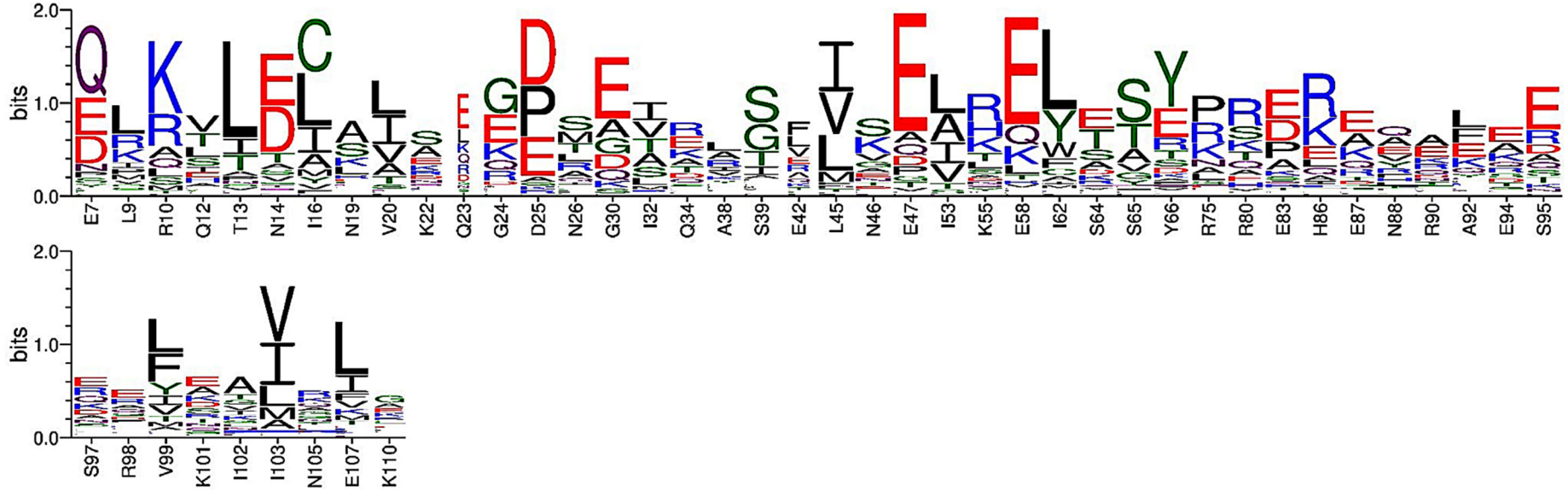


Figure S1: Sequence logo of LmrR multisequence alignment (MSA) made with the Weblogo software.[22,23] Only the positions that were identified as key dynamical hotpots are shown. Notice that F54 does not appear here, as it was added to the library since it was identified as a consensus mutation.

Identification of the improved variants

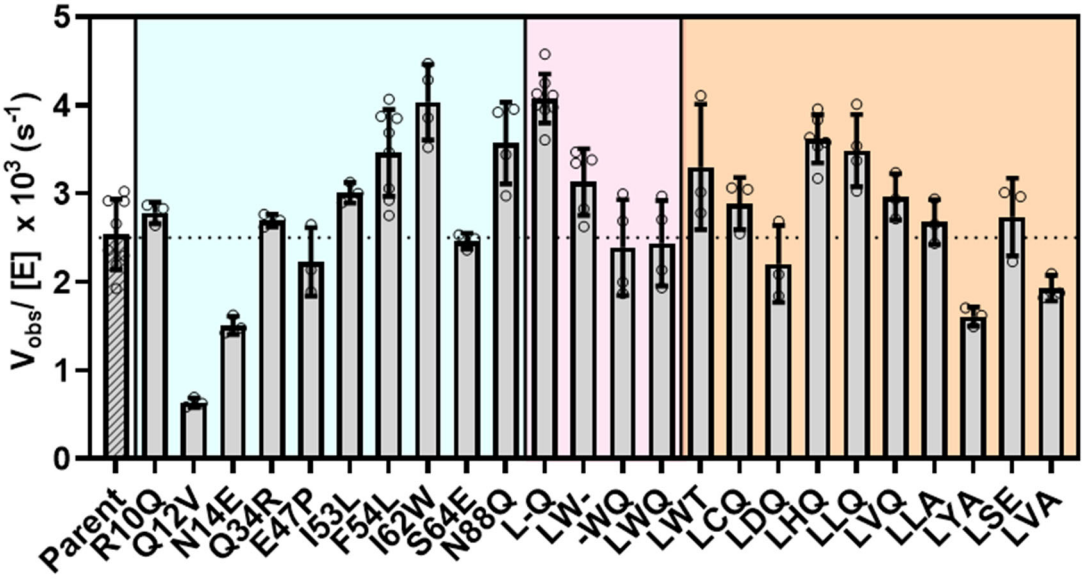


Figure S2: Rate of hydrazone formation measured with purified protein variants. Cyan box: point mutants isolated from the initial library of distal mutants, pink box: recombination of the three best performing single mutant hits (F54L, I62W and N88Q), orange box: best performing variants identified from the combinatorial library between I62 and N88. Rate was measured with 50 µM NBD-H, 5 mM 4-HBA, 2.5 µM of, buffer R pH 7.5, 5% DMF, 25 ^{o}C. The parent is LmrR_RMH and is shown with diagonal stripes and its rate is also reported with a dashed line across the plot. Error bars represent the standard deviation of at least three measurements.

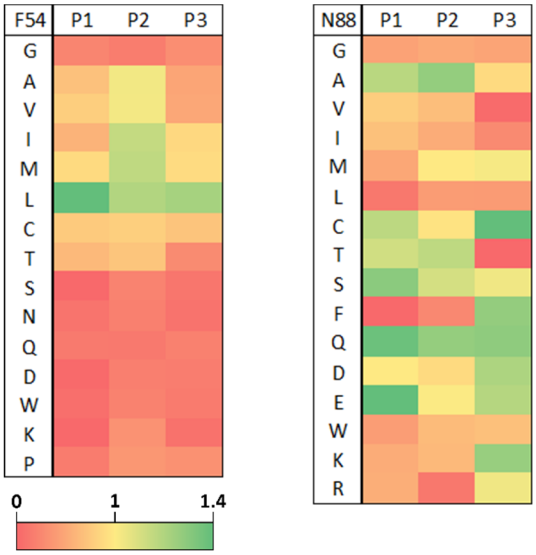


Figure S3: Partial site saturation of positions F54 and N88. The values represent the activity observed in cell lysate relative to the parent LmrR_RMH. Each column represents an independent measurement (i.e. P1, P2, P3).

Combinatorial library I62+N88

Table S2: Mutants showing more than 1.2-fold improved activity from LmrR_RMH_F54L in cell lysate screening of the combinatorial library at positions I62 and N88.

TIMES FOUND	MUTANT		
	I62	**N88**	**Name**
2	V	Q	LVQ
2	W	T	LWT
1	L	Q	LLQ
1	S	E	LSE
1	Y	A	LYA
1	N	Q	LNQ
1	C	Q	LCQ
1	R	Q	LRQ
4	H	Q	LHQ
1	L	A	LLA
1	D	Q	LDQ
1	W	R	LWR
1	V	A	LVA

Kinetic characterization

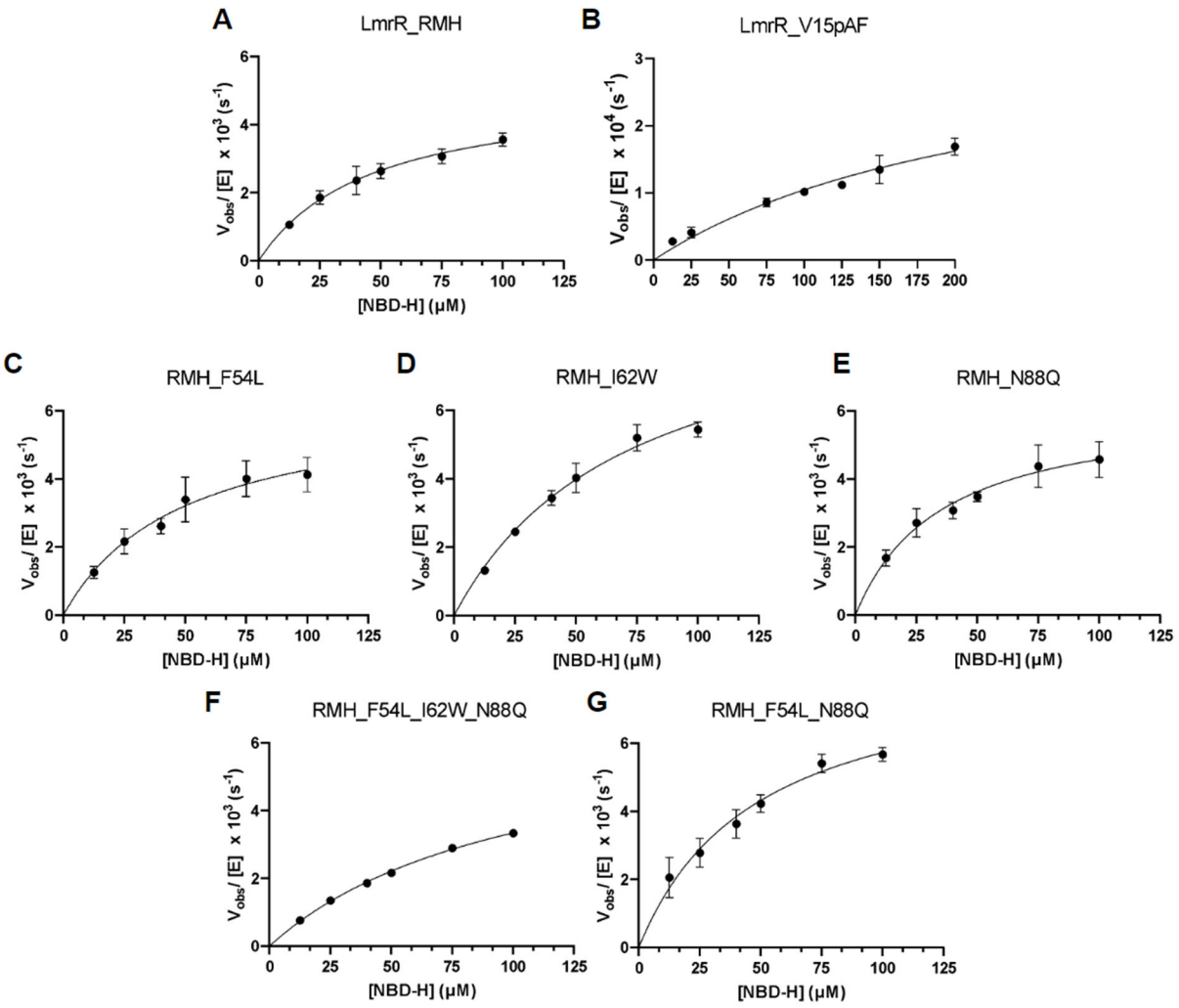


Figure S4: Apparent kinetic characterization of **A**) LmrR_RMH, **B)** LmrR_V15pAF, **C)** LmrR_RMH_F54L, **D)** LmrR_RMH_I62W, **E)** LmrR_RMH_N88Q, **F)** LmrR_RMH_LWQ, **G)** LmrR_RMH_L-Q. Error bars represent the standard deviation of at least two replicates from two independent batches of purified enzymes, for a total of at least four measurements for each point. Missing error bars indicate error too low to be depicted. The apparent kinetics of LmrR_V15pAF was determined with two replicates from one batch of enzyme. The reactions were carried out at fixed concentration of 4-HBA (5 mM) and different concentrations of NBD-H, 2.5 μM of enzyme (5 μM for LmrR_V15pAF), buffer R pH 7.5, 5% DMF, 25 ^{o}C.

Table S3: Apparent kinetic parameters for LmrR_pAF variants used in this study. Standard conditions used are fixed 4-HBA concentration (5 mM) different concentrations of NBD-H from 12.5 µM to 100 µM, 2.5 µM of enzyme, buffer R pH 7.5, 5% DMF, 25 °C. The standard deviation of K_M and k_{cat} is reported in parenthesis and is the result of at least four experiments composed of two technical and two biological duplicates.

Variant	$K_{M,\,app}$ **(µM)**	$k_{cat,\,app}$ **x** 10^{-3} **(s^{-1})**	$(k_{cat}/K_M)_{app}$ **($M^{-1}s^{-1}$)**	$(k_{cat}/K_M)_{app}$ **relative to LmrR_RMH[a]**	$(k_{cat}/K_M)_{app}$ **relative to LmrR_pAF[ab]**
LmrR_pAF*	243.5 (±80.8)	0.36 (±0.08)	1.48	0.01	1
LmrR_RMH	46.1 (±6.8)	5.11 (±0.34)	111	1	75
LmrR_RMH_LWQ	104.0 (±5.4)	7.86 (±0.37)	140	1.3	95

[a]The comparisons with LmrR_RMH and LmrR_V15pAF were made based on newly measured values. [b]The newly measured activity for LmrR_V15pAF was 10-fold higher from what was originally reported. This variation is mostly due to the relatively low-rate acceleration achieved by the non-engineered LmrR_V15pAF, which introduces a higher error rate during measurement. Standard conditions were modified as follows: 5 µM of enzyme, concentration of NBD-H ranged from 12.5 µM to 200 µM to ensure saturation. The standard deviation reported is the result of a technical duplicate.*

Analysis of thermostability

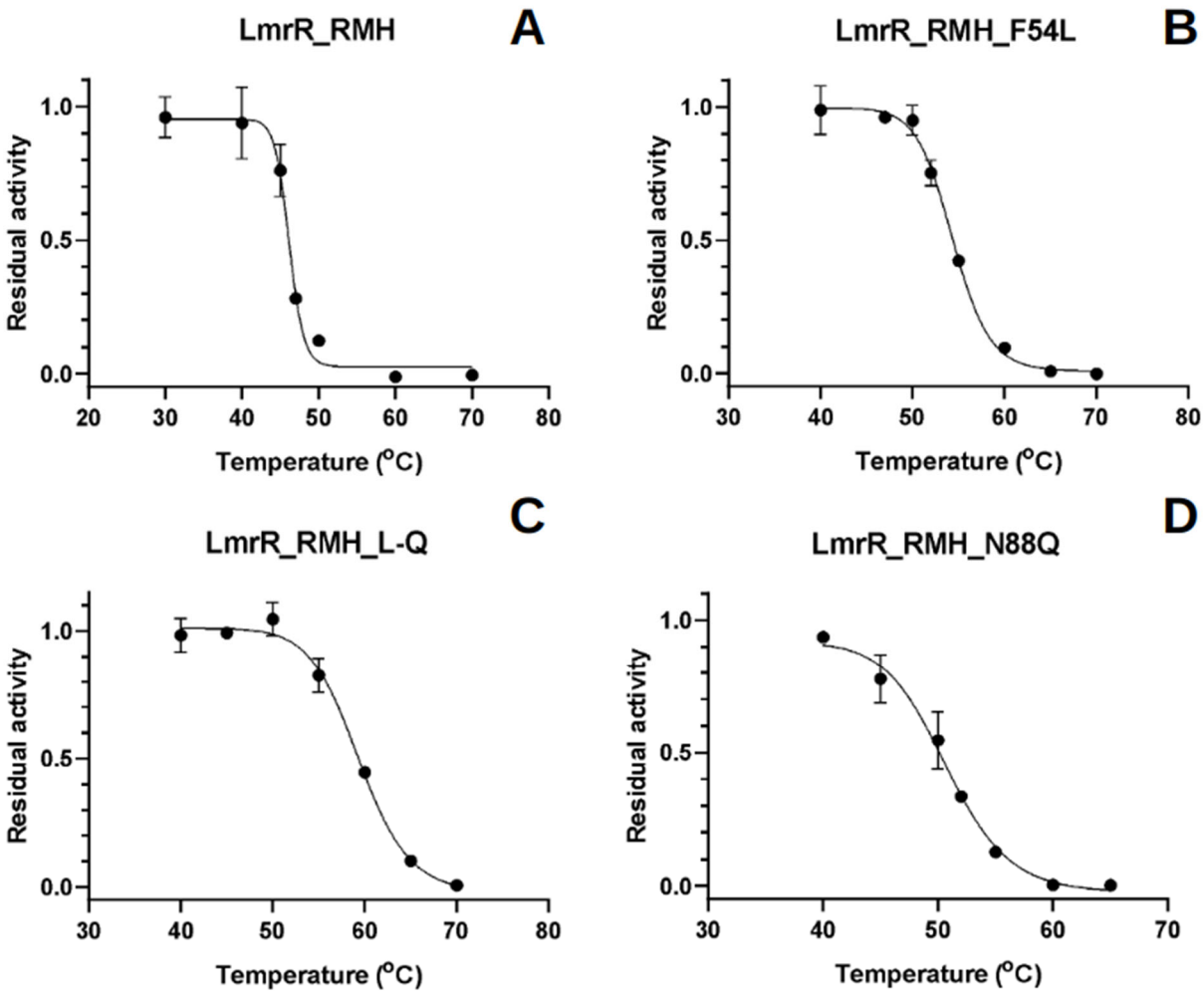


Figure S5: Operational thermostability measurements of (**A**) LmrR_RMH, (**B**) LmrR_RMH_F54L, (**C**) LmrR_RMH_L-Q, and (**D**) LmrR_RMH_N88Q upon incubation at different temperatures. Each data point represents the average of three independent experiments for which the standard deviation is reported. The fitted line is a standard 4PL sigmoidal curve.

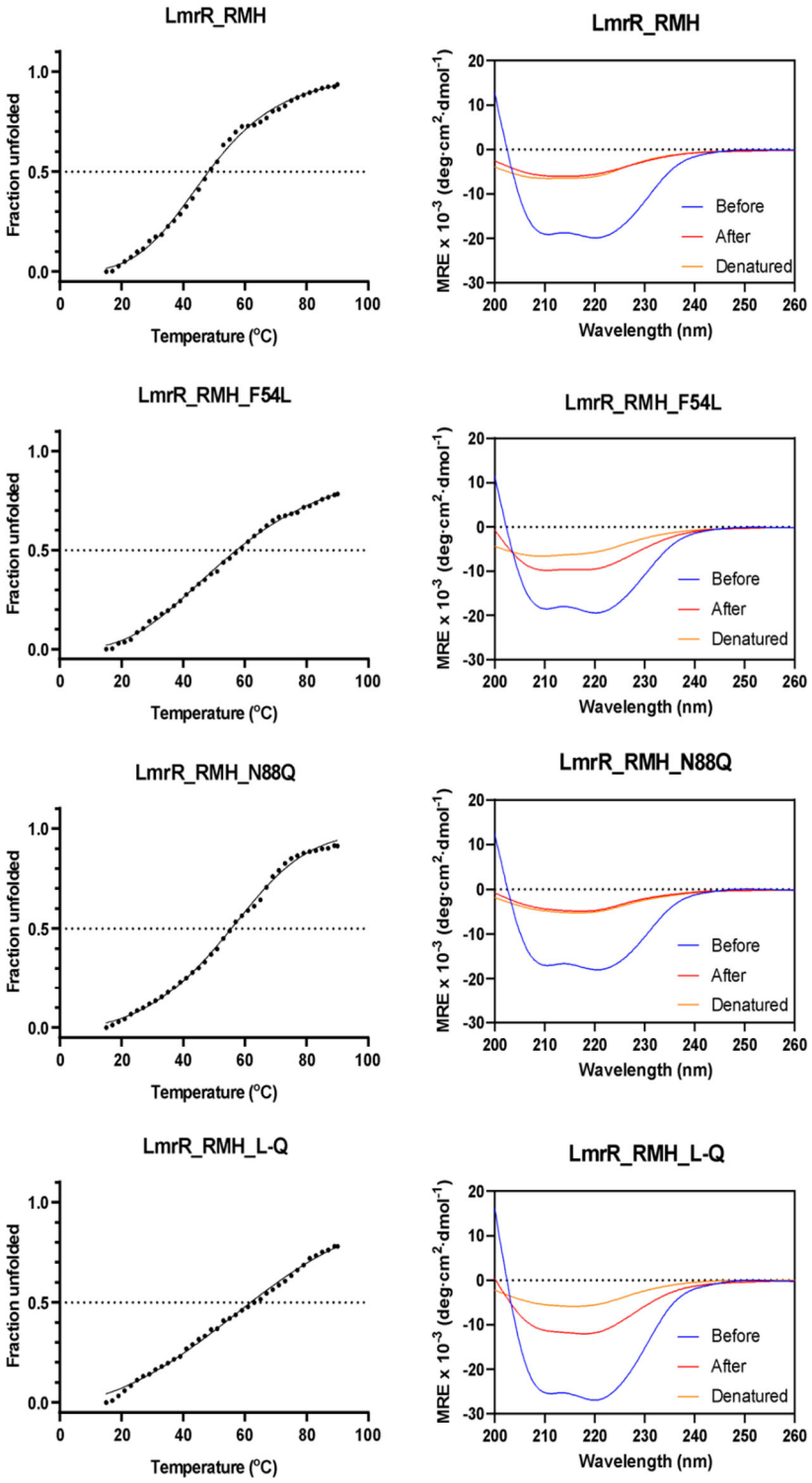


Figure S6: Determination of structural melting temperature T_m based on circular dichroism spectroscopy. The left panels represent the thermal unfolding of the secondary structure of LmrR_RMH and mutants F54L, N88Q and L-Q followed at 222 nm. Dashed line represents the position of the T_{50}^{10} calculated. The right panels depict the CD spectrum before (blue) and after (red) the thermal treatment. As a reference, the CD spectrum of each fully denatured mutant (yellow) is reported.

Table S4: Structural (T_m) and operational (T_{50}^{10}) thermal stability parameters of the mutants F54L, N88Q and F54L_N88Q (L-Q) and their increase compared to the parent. In brackets, the standard deviation of three independent measurements.

Mutant	**T_m**	**ΔT_m**	T_{50}^{10}	**ΔT_{50}^{10}**
LmrR_RMH	48.2 (±0.6)	0	46.2 (±0.4)	0
LmrR_RMH_F54L	57.3 (±1.1)	9.1	54.4 (±0.5)	8.2
LmrR_RMH_N88Q	55.7 (±0.6)	7.5	50.8 (±0.8)	4.6
LmrR_RMH_L-Q	61.9 (±1.6)	13.7	59.4 (±0.8)	13.2

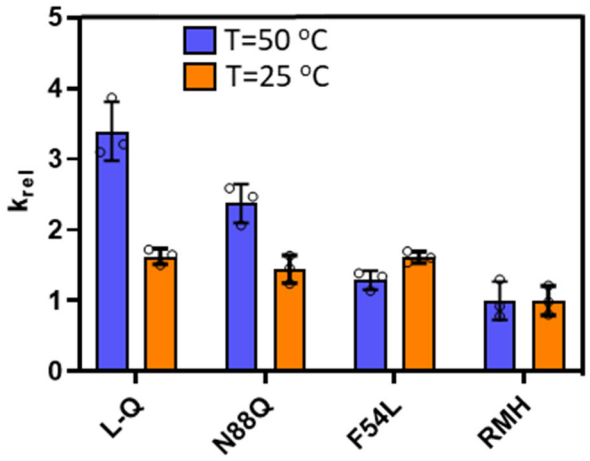


Figure S7: Relative catalytic activity of LmrR_RMH (RMH), LmrR_RMH_F54L (F54L), LmrR_RMH_N88Q (N88Q) and LmrR_RMH_L-Q (L-Q) at 50 °C (blue) and 25 °C (orange). Error bars represent the standard deviations of three independent replicates, each shown with empty circles.

Molecular dynamics simulations

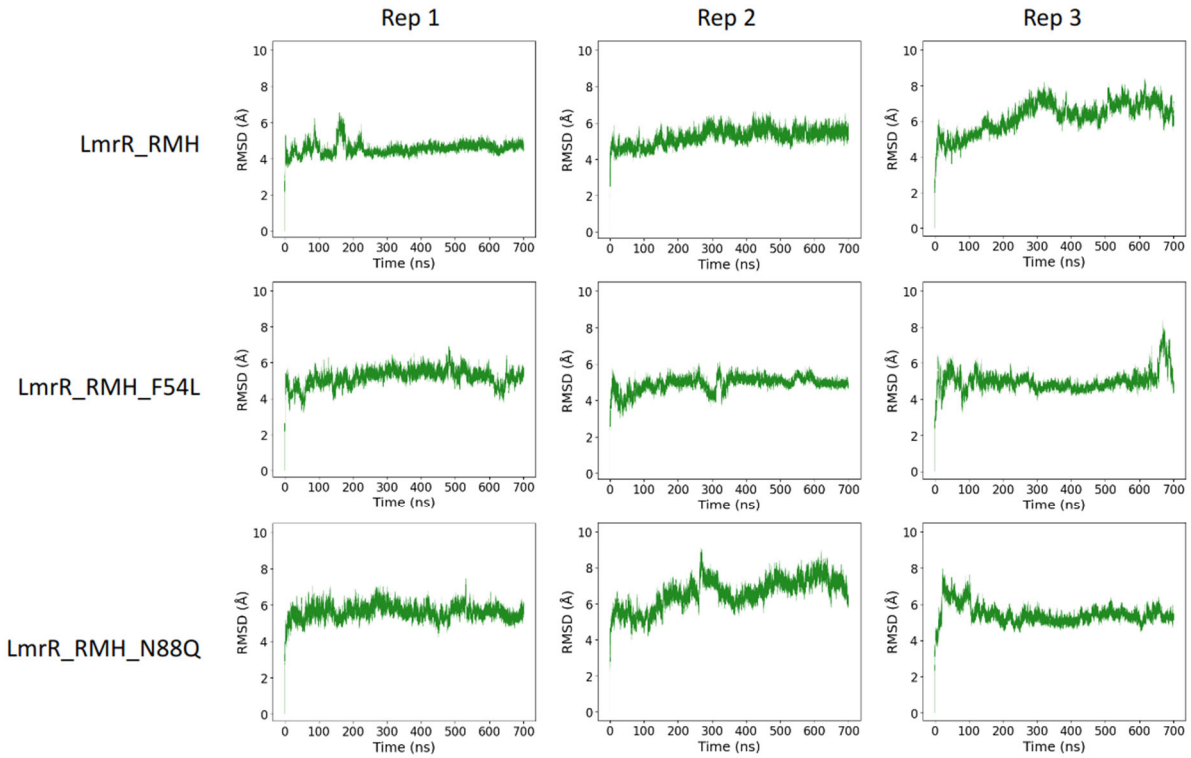


Figure S8: Root-Mean Square Deviation (RMSD) values of the parent LmrR_RMH (top row), LmrR_RMH_F54L (middle row) and LmrR_RMH_N88Q (bottom row) along the MD trajectories in each of the three replicas with reference to the first frame of the trajectory.

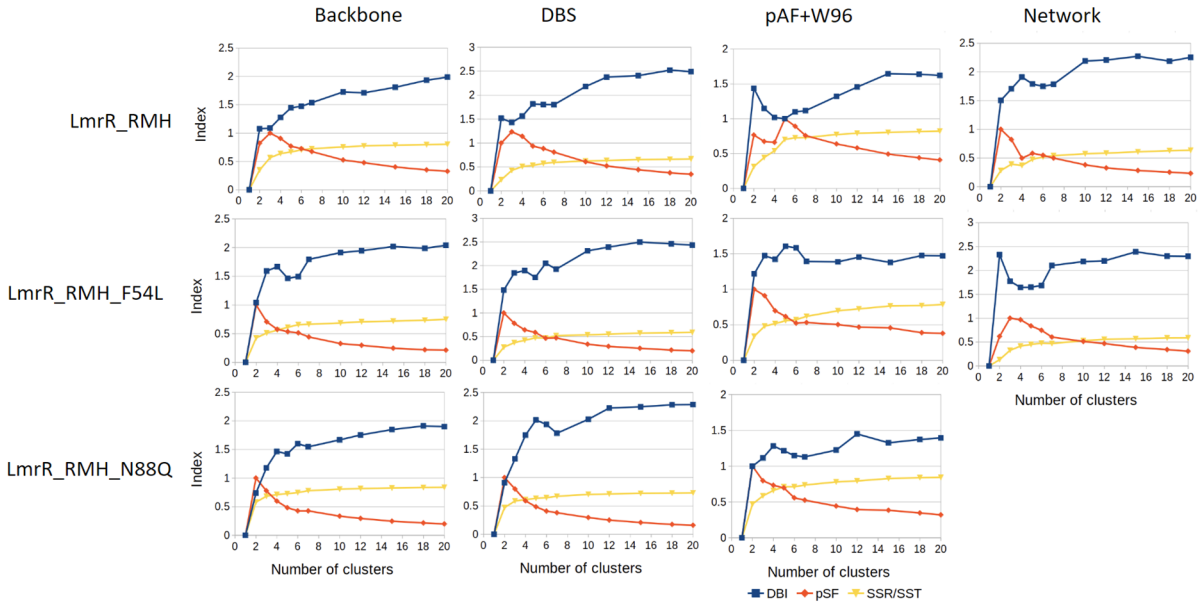


Figure S9: Analysis of the clustering parameters DBI (blue), pSF (red) and the ratio SSR/SST (yellow) trends over several numbers of clusters based on different selection of residues (see Computational methods) for LmrR_RMH, and the mutants F54L and N88Q. The value of pSF reported is normalised to ensure proper fitting of the curve in the plot. DBS: DNA-binding site. DBS:

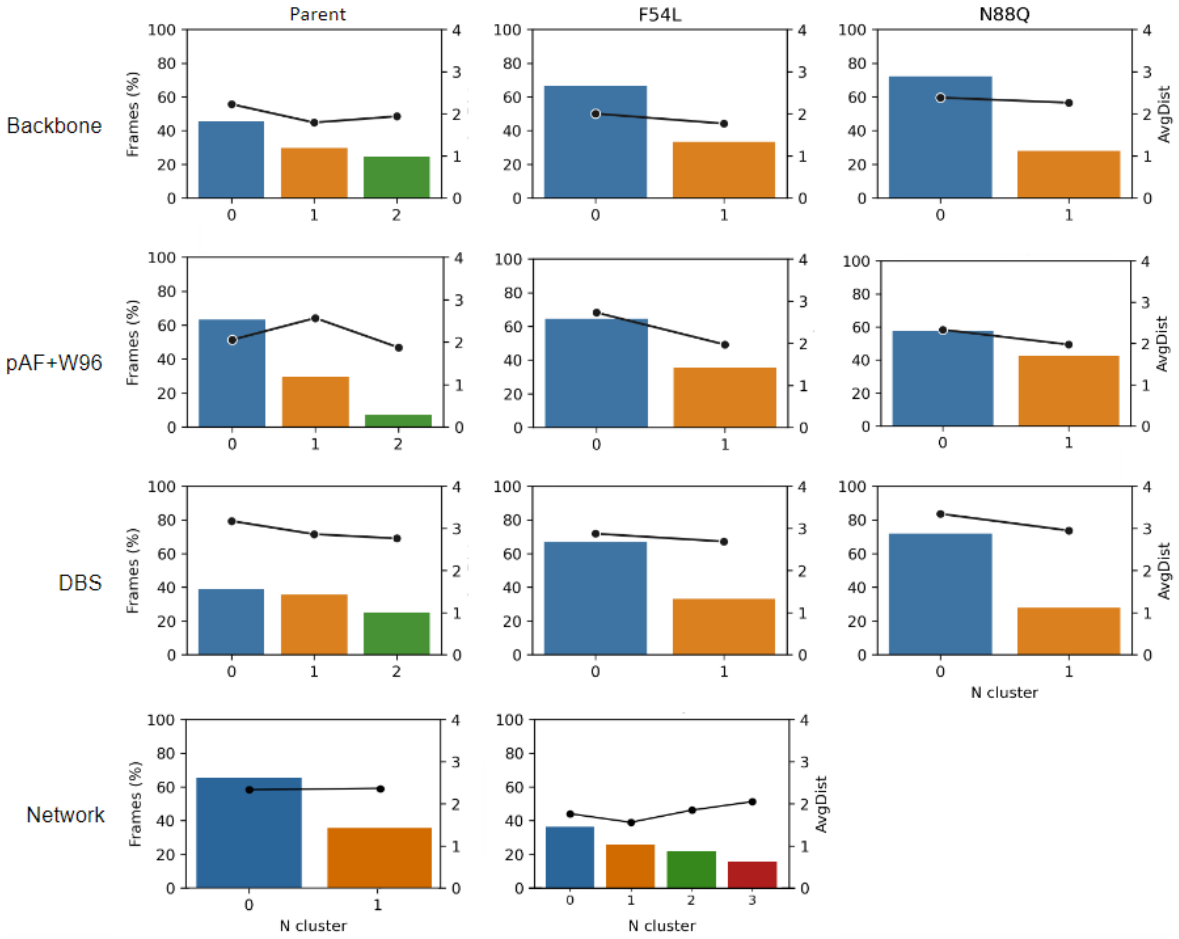


Figure S10: Representation of the number of frames (bars) populating each of the clusters selected for inspection as well as the average within-cluster deviation (line). Clustering was performed based on different selections of residues (see Computational methods).

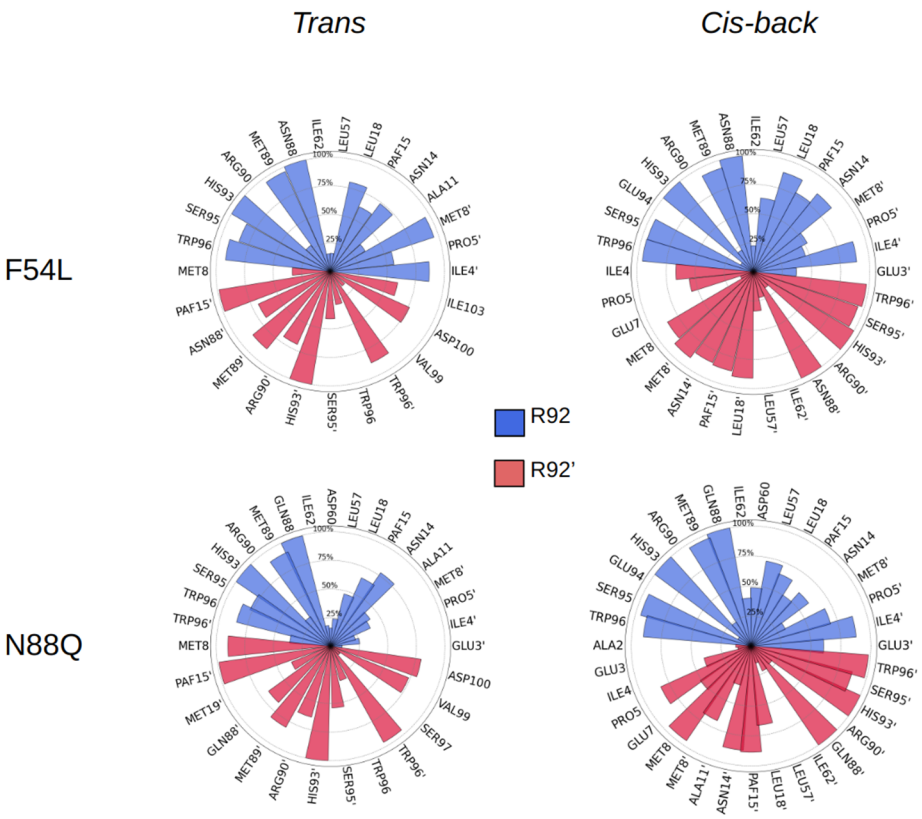


Figure S11: Residues in contact with R92 (blue) or R92' (red) are reported together with their frequency of interactions in each of the states identified. A cut-off of 10 % contact frequency was used.

Table S5: Number of frames identified in each protein for each of the three states.

Mutant	Trans state		Cis back state		Cis front state	
	Frames	Of the total trajectory	Frames	Of the total trajectory	Frames	Of the total trajectory
LmrR_RMH	1870	8.8%	10711	50.5%	5584	26.3%
F54L	7706	36.3%	11717	55.2%	0	0.0%
N88Q	8517	40.1%	8408	39.4%	36	0.0%

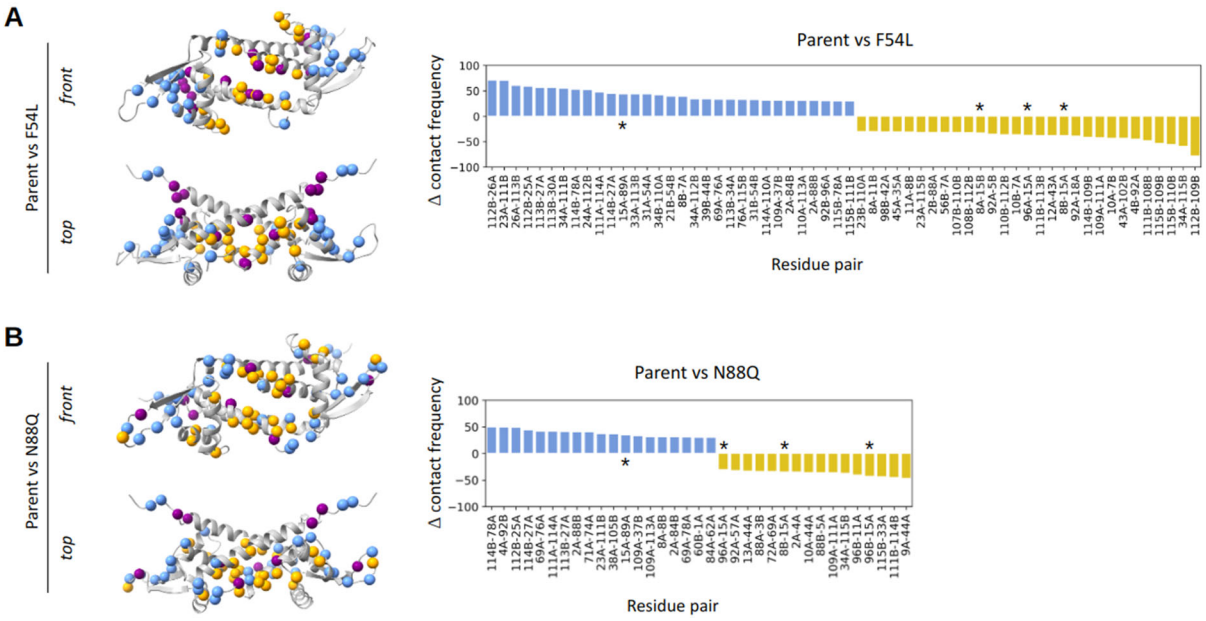


Figure S12: The relative frequency of contacts between the parent LmrR_RMH and the F54L (**A**) and N88Q (**B**) mutants using a cutoff of 30% is represented. On the left, top and front view of LmrR_RMH structure. The residues showing a high contact frequency are depicted in balls, in yellow when this frequency was higher for the mutant LmrR_RMH_F54L, in blue for the parent and in purple if the residues showed a high number of contacts in both the parent and the mutant. On the right, the contacting residues are reported. Consistently, the contacts that are more frequent in the LmrR_RMH_F54L are shown in yellow, and the contacts which are more frequent in the parent are shown in blue. *Contacts involving the non-canonical residue pAF.

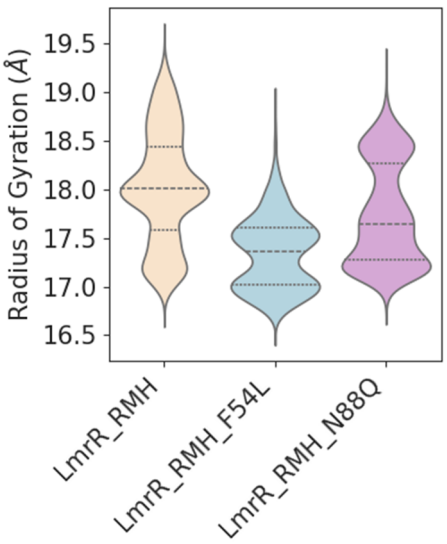


Figure S13: Radius of gyration of all the alpha helices of the three protein variants LmrR_RMH, LmrR_RMH_F54L and LmrR_RMH_N88Q.

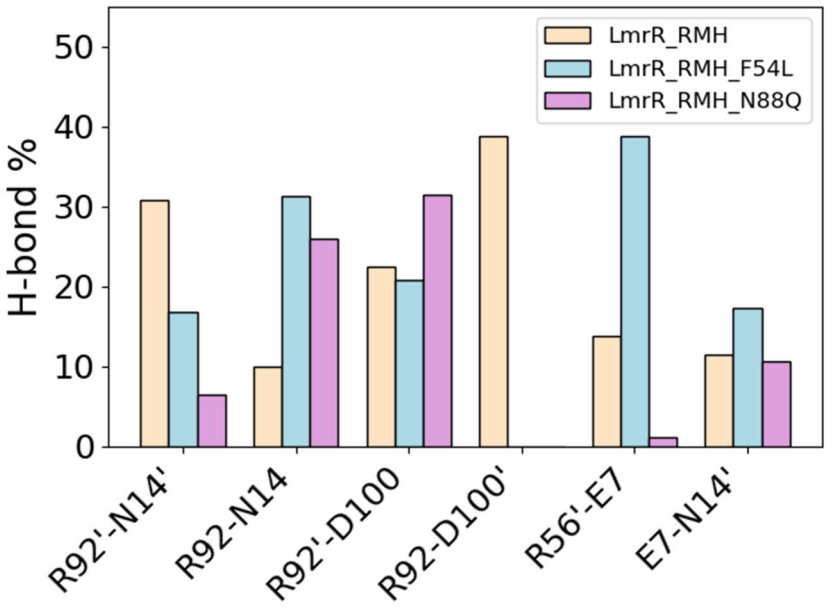
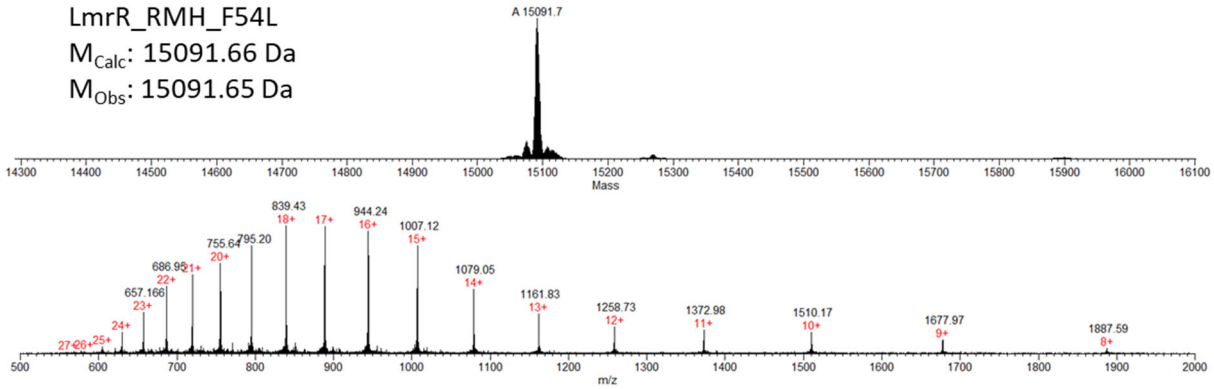
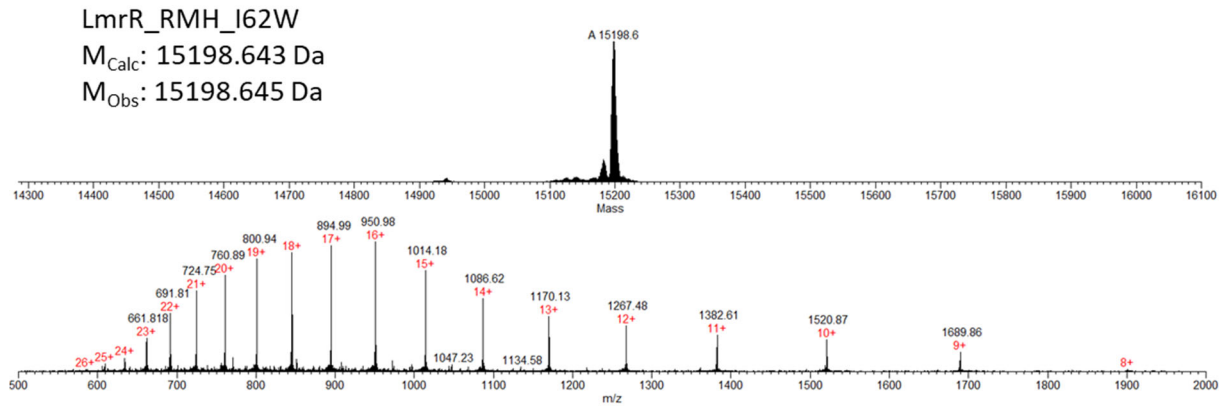


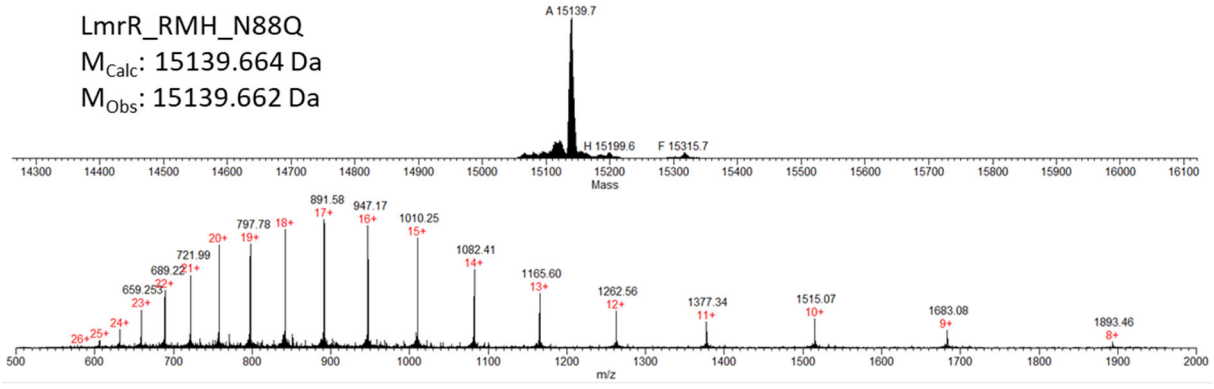
Figure S14: Frequency of hydrogen bond formation along the full trajectory (2.1 μs) for various pairs of residues for the parent, LmrR_RMH, and the mutants LmrR_RMH_F54L and LmrR_RMH_N88Q. Notably, the interactions observed were not symmetrical between the monomers. No interactions between R92 and D100' were observed in both the mutants.

Mass spectra of purified proteins

HR-MS spectra



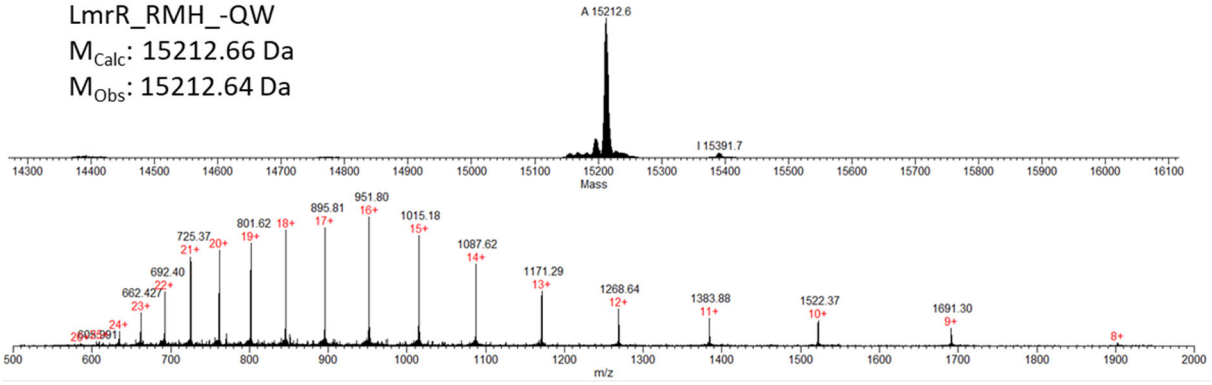




LmrR_RMH_-QW

M_{Calc}: 15212.66 Da

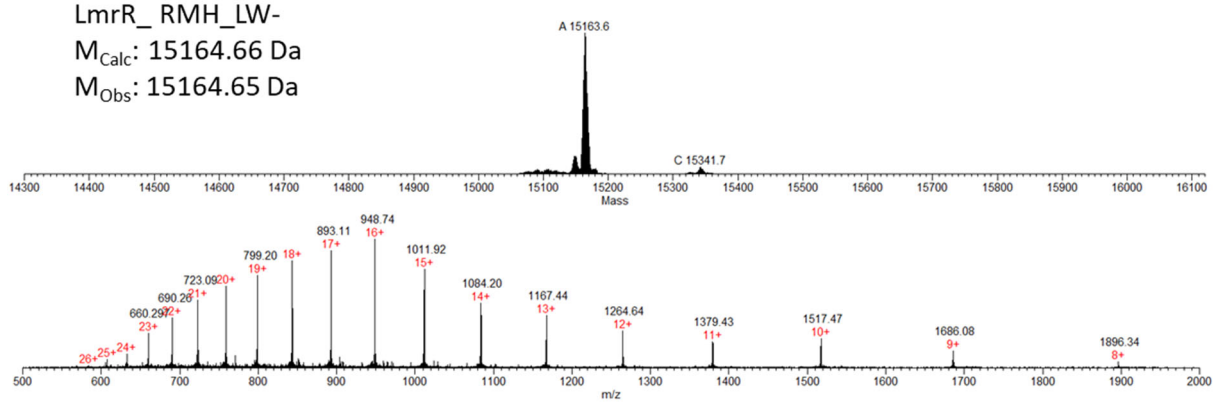
M_{Obs}: 15212.64 Da



LmrR_ RMH_LW-

M_{Calc}: 15164.66 Da

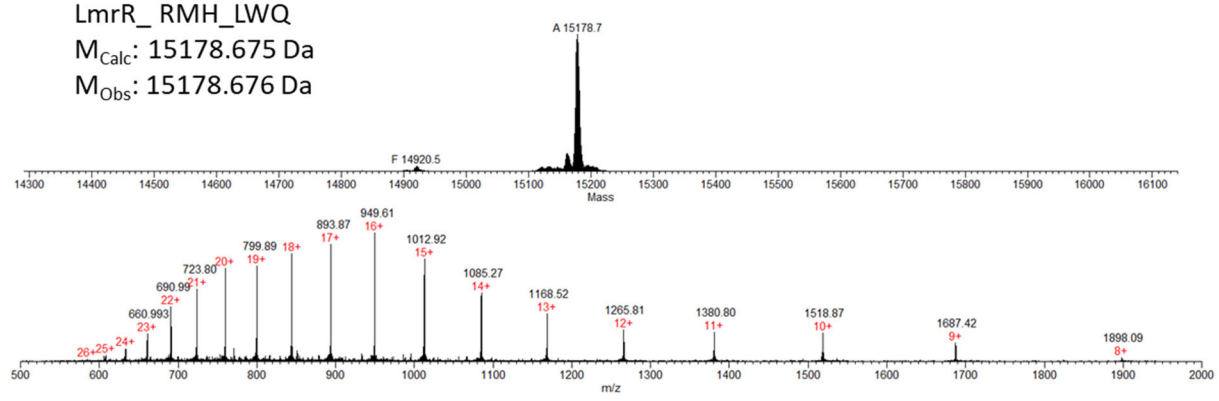
M_{Obs}: 15164.65 Da



LmrR_ RMH_LWQ

M_{Calc}: 15178.675 Da

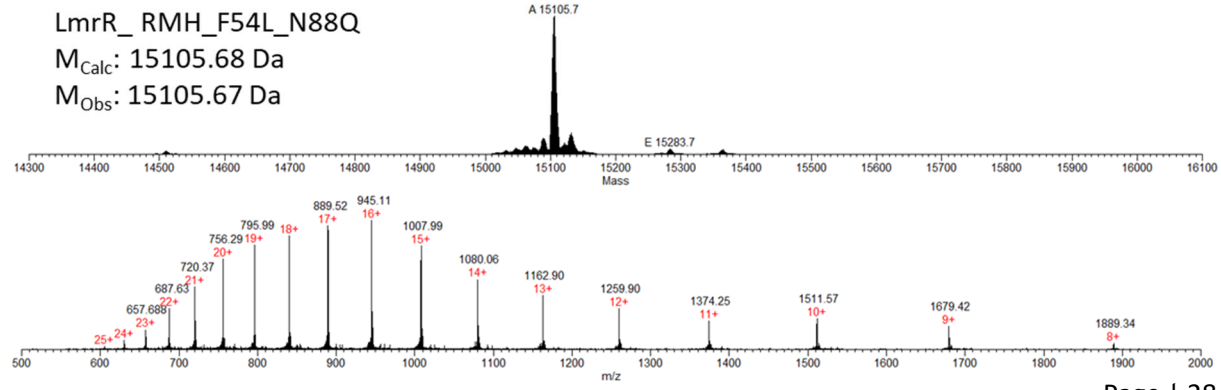
M_{Obs}: 15178.676 Da



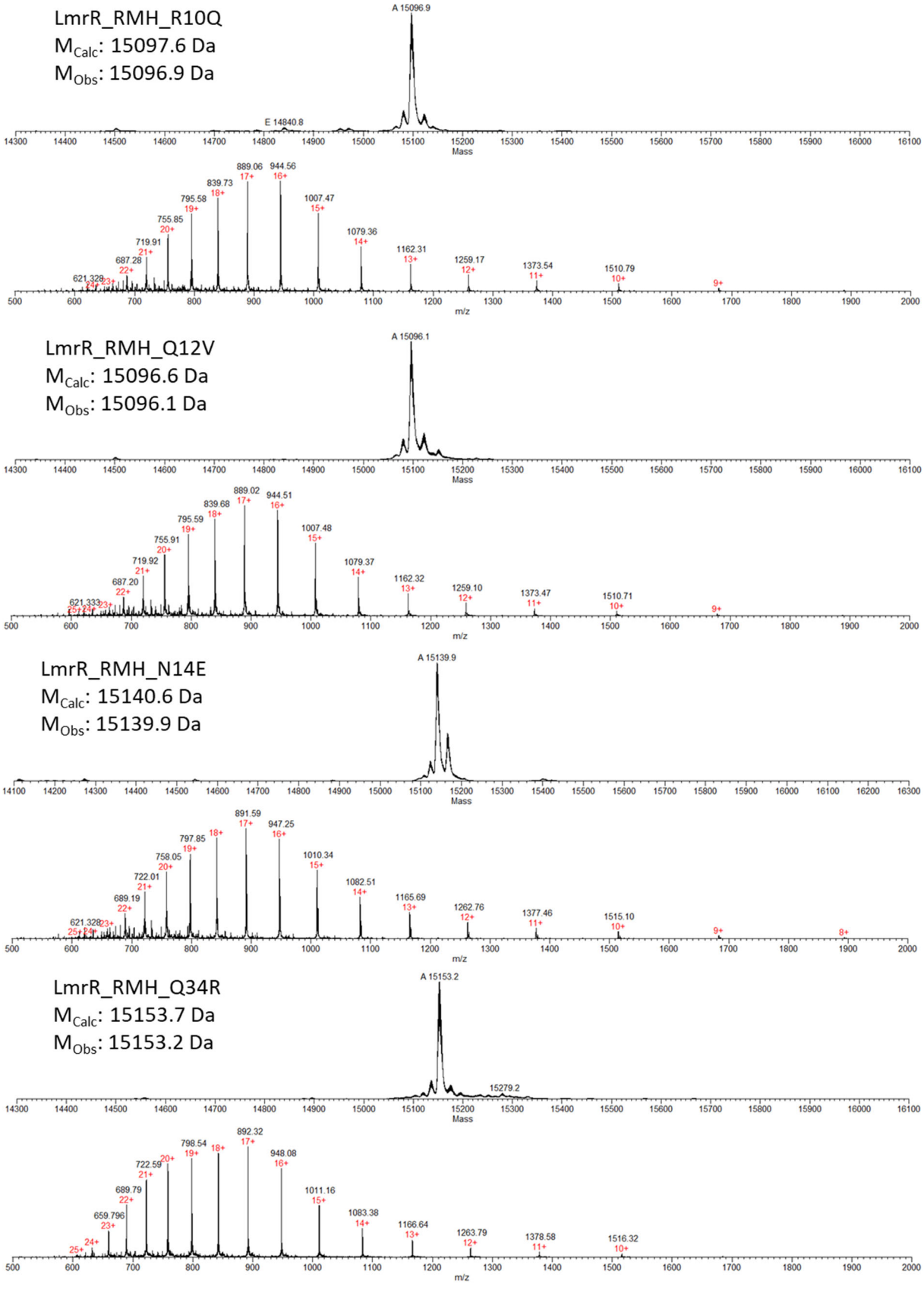
LmrR_ RMH_F54L_N88Q

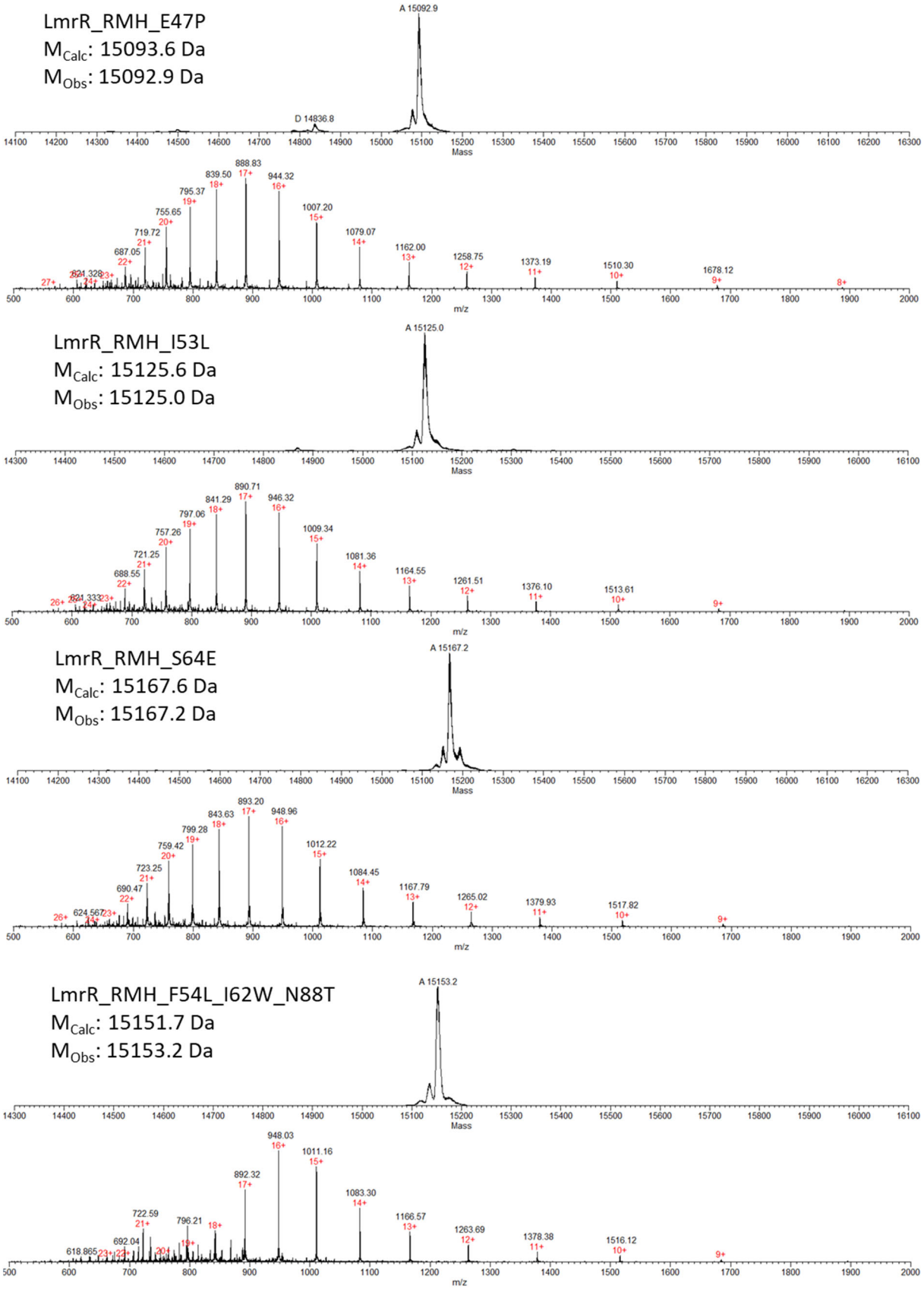
M_{Calc}: 15105.68 Da

M_{Obs}: 15105.67 Da



Low resolution-MS spectra

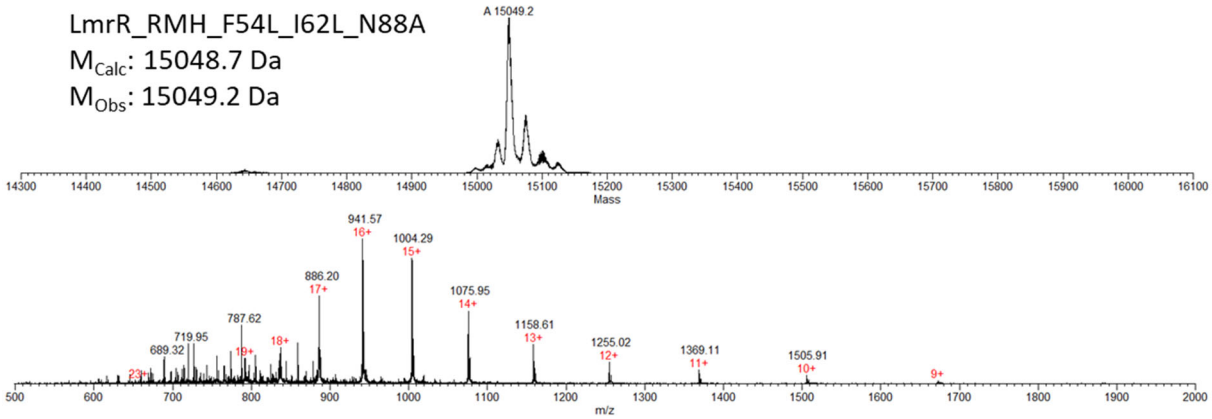



LmrR_RMH_E47P
M_{Calc}: 15093.6 Da
M_{Obs}: 15092.9 Da
A 15092.9
D 14836.8
Mass
888.83 17+
839.50 18+
944.32 16+
795.37 19+
1007.20 15+
755.65 20+
1079.07 14+
719.72 21+
1162.00 13+
687.05 22+
1258.75 12+
1373.19 11+
1510.30 10+
1678.12 9+
m/z
LmrR_RMH_I53L
M_{Calc}: 15125.6 Da
M_{Obs}: 15125.0 Da
A 15125.0
Mass
890.71 17+
841.29 18+
946.32 16+
797.06 19+
1009.34 15+
757.26 20+
1081.36 14+
721.25 21+
1164.55 13+
688.55 22+
1261.51 12+
1376.10 11+
1513.61 10+
m/z
LmrR_RMH_S64E
M_{Calc}: 15167.6 Da
M_{Obs}: 15167.2 Da
A 15167.2
Mass
893.20 17+
843.63 18+
948.96 16+
799.28 19+
1012.22 15+
759.42 20+
1084.45 14+
723.25 21+
1167.79 13+
690.47 22+
1265.02 12+
1379.93 11+
1517.82 10+
m/z
LmrR_RMH_F54L_I62W_N88T
M_{Calc}: 15151.7 Da
M_{Obs}: 15153.2 Da
A 15153.2
Mass
948.03 16+
1011.16 15+
892.32 17+
1083.30 14+
722.59 21+
796.21 18+
1166.57 13+
1263.69 12+
692.04 22+
1378.38 11+
1516.12 10+
m/z

LmrR_RMH_F54L_I62L_N88A

M_{Calc}: 15048.7 Da

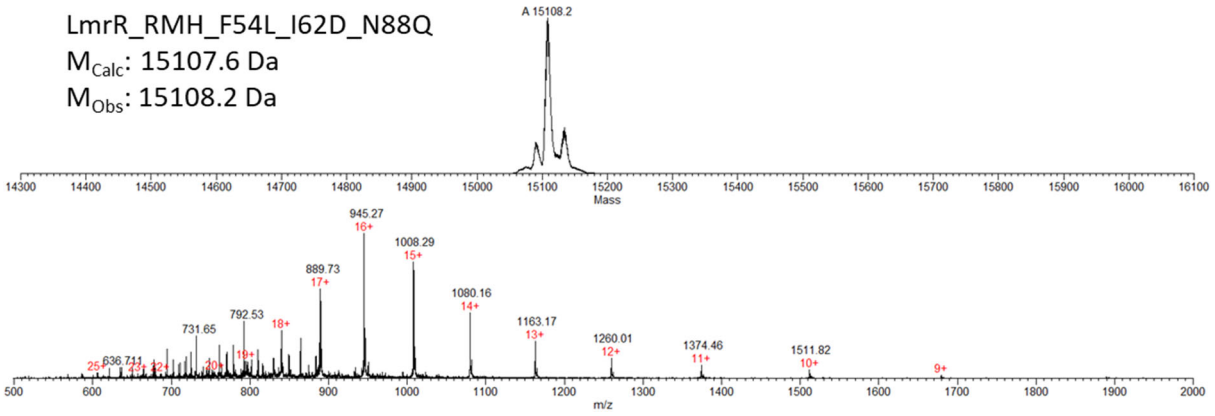
M_{Obs}: 15049.2 Da



LmrR_RMH_F54L_I62D_N88Q

M_{Calc}: 15107.6 Da

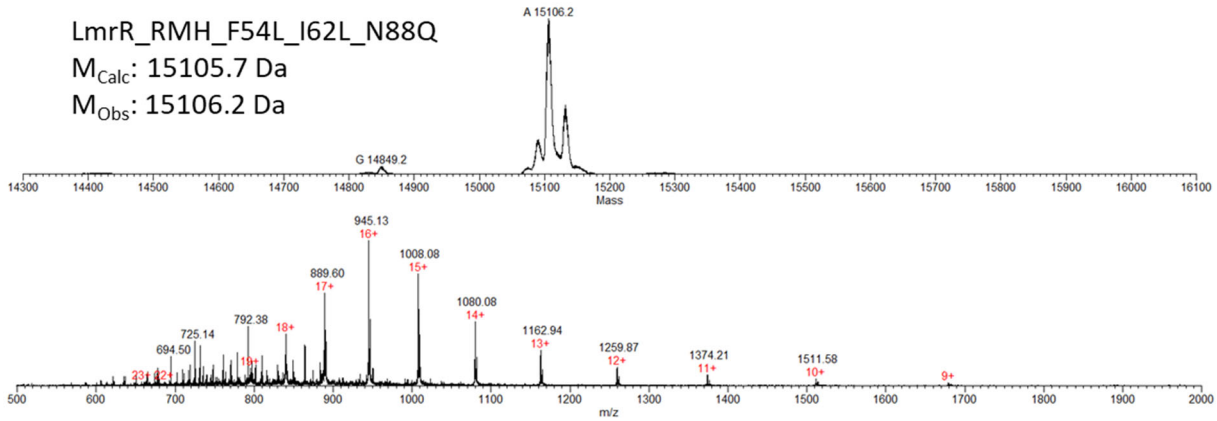
M_{Obs}: 15108.2 Da



LmrR_RMH_F54L_I62L_N88Q

M_{Calc}: 15105.7 Da

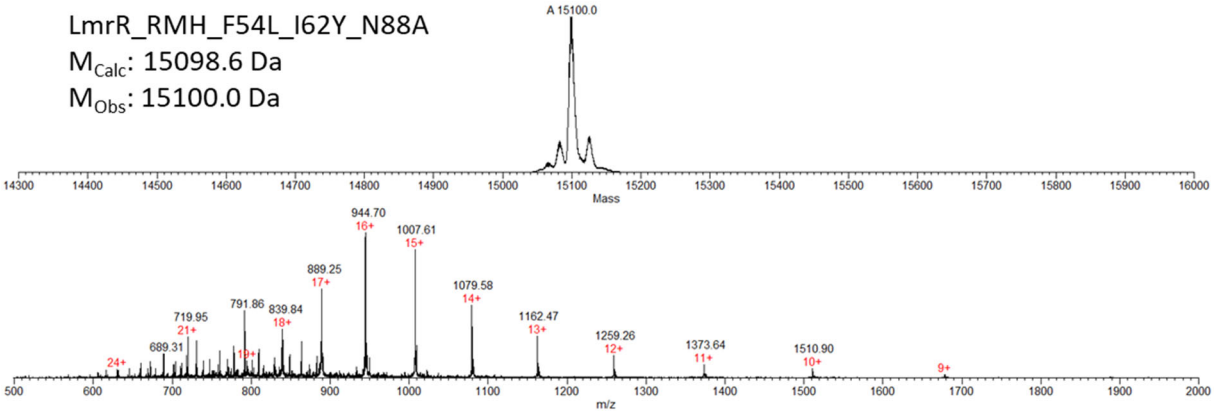
M_{Obs}: 15106.2 Da



LmrR_RMH_F54L_I62Y_N88A

M_{Calc}: 15098.6 Da

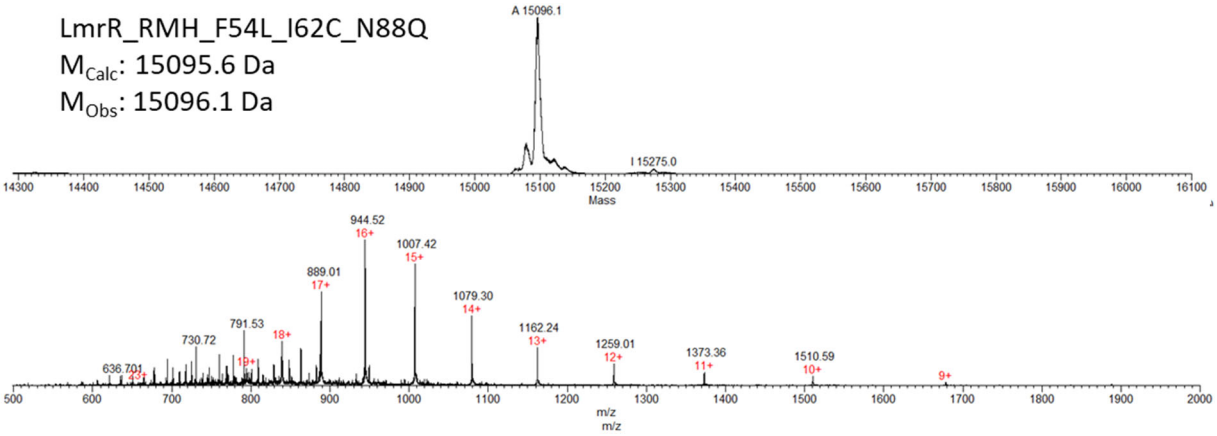
M_{Obs}: 15100.0 Da



LmrR_RMH_F54L_I62C_N88Q

M_{Calc}: 15095.6 Da

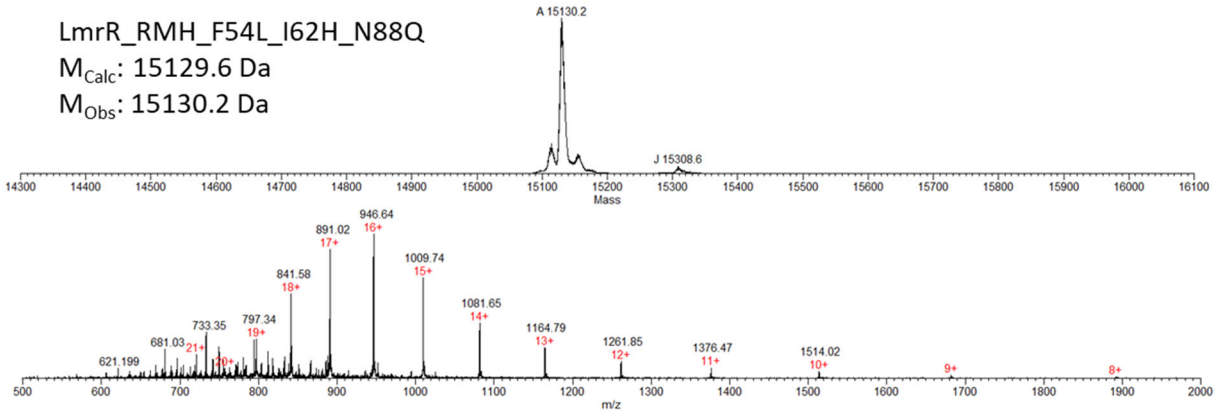
M_{Obs}: 15096.1 Da



LmrR_RMH_F54L_I62H_N88Q

M_{Calc}: 15129.6 Da

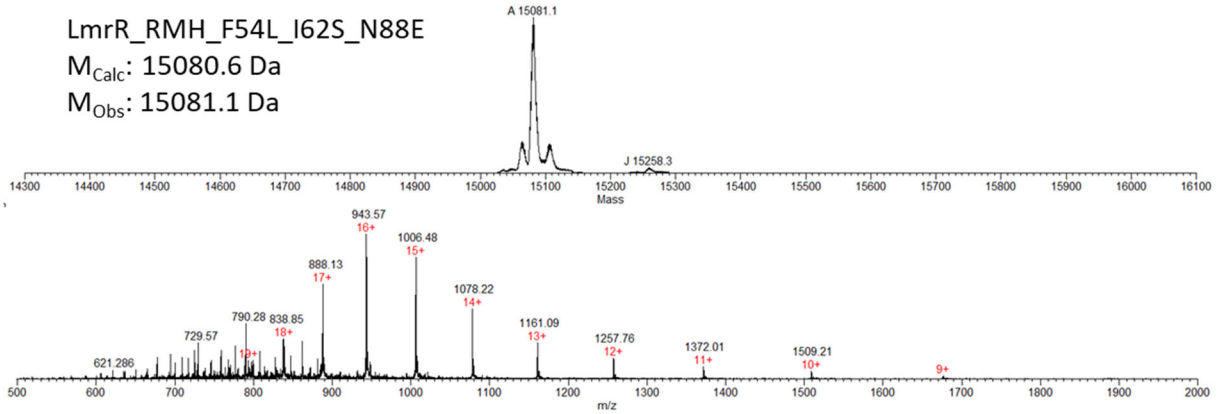
M_{Obs}: 15130.2 Da



LmrR_RMH_F54L_I62S_N88E

M_{Calc}: 15080.6 Da

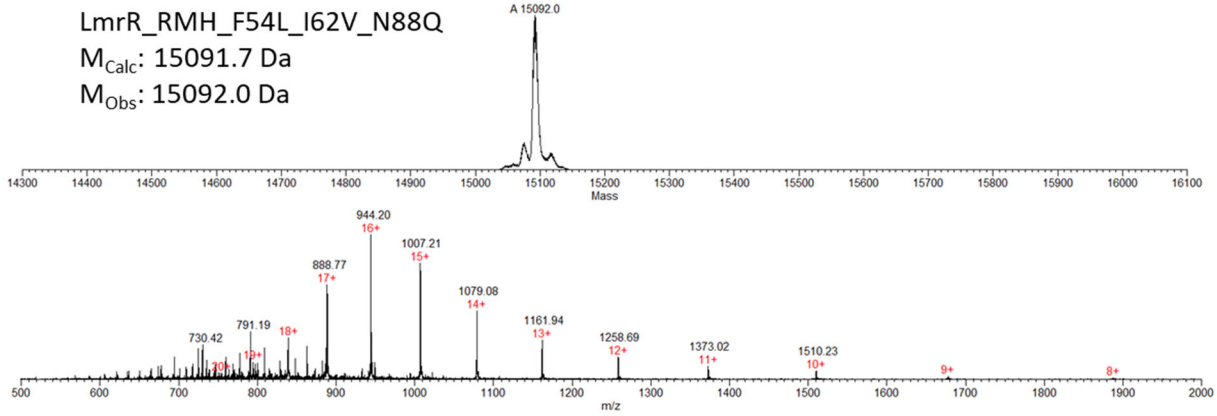
M_{Obs}: 15081.1 Da



LmrR_RMH_F54L_I62V_N88Q

M_{Calc}: 15091.7 Da

M_{Obs}: 15092.0 Da



List of primers used

Table S6: DNA primers employed to construct the distal point mutation library. Underlined is the codon that introduces the mutation desired.

Primer Name	Sequence 3' 5'	Primer Name	Sequence 3' 5'
E7D_fwd	CGAAAgatATGCTGCGTGCTCAAACC	E7D_rv	GCAGCATatcTTTCGGGATTTCGGC
L9K_fwd	AGAAATGaaaCGTGCTCAAACCAATTAGATCC	L9K_rv	GAGCACGtttCATTTCTTTCGGGATTTCGG
L9R_fwd	GAAATGcgtCGTGCTCAAACCAATTAGATCC	L9R_rv	GAGCACGacgCATTTCTTTCGGGATTTCGG
R10Q_fwd	AATGCTGcagGCTCAAACCAATTAGATCCTGC	R10Q_rv	TTTGAGCctgCAGCATTTCTTTCGGGAT
R10K_fwd	AATGCTGaaaGCTCAAACCAATTAGATCCTGCTG	R10K_rv	GGTTTGAGCtttCAGCATTTCTTTCGGGATTTCG
R10A_fwd	AATGCTGgcgGCTCAAACCAATTAGATCCTGCTG	R10A_rv	GTTTGAGCcgcCAGCATTTCTTTCGGGATTTCGG
Q12V_fwd	GCGTGCTgtgACCAATTAGATCCTGCTGATGG	Q12V_rv	TAATTGGTcacAGCACGCAGCATTTCTTTCGG
Q12E_fwd	GCGTGCTgaaACCAATTAGATCCTGC	Q12E_rv	ATTGGTttcAGCACGCAGCATTTC
T13I_fwd	GCTCAAatcAATTAGATCCTGCTGATGG	T13I_rv	TCTAATTgatTTGAGCACGCAGCATTTCTTTC
T13L_fwd	GCGTGCTCAActgAATTAGATCCTGCTGATGGTC	T13L_rv	GATCTAATTcagTTGAGCACGCAGCATTTCTTTCG
N14E_fwd	CTCAAACCgaaTAGATCCTGCTGATGGTCC	N14E_rv	GCAGGATCTAttcGGTTTGAGCACGCAGCA
I16M_fwd	CCAATTAGatgCTGCTGATGGTCCTGAAAC	I16M_rv	CCATCAGCAGcatCTAATTGGTTTGAGCACG
I16C_fwd	CCAATTAGtgcCTGCTGATGGTCCTGAAACAAGG	I16C_rv	CCATCAGCAGgcaCTAATTGGTTTGAGCACG
I16L_fwd	CCAATTAGctgCTGCTGATGGTCCTGAAACAAGG	I16L_rv	CCATCAGCAGcagCTAATTGGTTTGAGCACG
V20I_fwd	CTGCTGATGattCTGAAACAAGGCGATAACTATG	V20I_rv	GCCTTGTTTCAGaatCATCAGCAGGATCTAATTGG
V20L_fwd	CTGCTGATGctgCTGAAACAAGGCGATAACTATG	V20L_rv	CTTGTTTCAGcagCATCAGCAGGATCTAATTGG
Q23L_fwd	GTCCTGAAActgGGCGATAACTATGTGTATGGC	Q23L_rv	GTTATCGCCcagTTTCAGGACCATCAGCAGG
G24E_fwd	CCTGAAACAAgaaGATAACTATGTGTATGGCA	G24E_rv	ACATAGTTATCttcTTGTTTCAGGACCATCAG

D25E_fwd	ACAAGGCgaaAACTATGTGTATGGCAT	D25E_rv	TACACATAGTTttcGCCTTGTTTCAGGACC
D25P_fwd	GAAACAAGGCcctAACTATGTGTATGGCATTATCA	D25P_rv	ACACATAGTTcggGCCTTGTTTCAGGACCATCAG
N26R_fwd	AGGCGATcgtTATGTGTATGGCATTATCAAACAGG	N26R_rv	ATACACATAacgATCGCCTTGTTTCAGGACCATCA
N26S_fwd	AGGCGATtctTATGTGTATGGCATTATCAAACAGG	N26S_rv	ATACACATAagaATCGCCTTGTTTCAGGACCATC
V28G_fwd	GCGATAACTATggtTATGGCATTATCAAACAGGTG	V28G_rv	ATGCCATAaccATAGTTATCGCCTTGTTTCAGG
G30S_fwd	CTATGTGTATagcATTATCAAACAGGTGAAAGAAG	G30S_rv	TGATAATgctATACACATAGTTATCGCCTTG
I32T_fwd	TGGCATTaccAAACAGGTGAAAGAAGCG	I32T_rv	CACCTGTTTggtAATGCCATACACATAG
I32V_fwd	TATGGCATTgtgAAACAGGTGAAAGAAGCGAG	I32V_rv	CACCTGTTTcacAATGCCATACACATAGTTATCG
Q34E_fwd	ATTATCAAAgaaGTGAAAGAAGCGAGCAACGG	Q34E_rv	CGCTTCTTTCACttcTTTGATAATGCCATACACA
Q34R_fwd	TTATCAAAcgtGTGAAAGAAGCGAGCAACGG	Q34R_rv	CGCTTCTTTCACacgTTTGATAATGCCATACAC
A38L_fwd	TGAAAGAActgAGCAACGGTGAAATGGAACTG	A38L_rv	ACCGTTGCTcagTTCTTTCACCTGTTTGATAATG
S39G_fwd	AAGAAGCGggcAACGGTGAA	S39G_rv	CCGTTgccCGCTTCTTTCACCTG
L45M_fwd	AATGGAAatgAATGAAGCCACCC	L45M_rv	TCATTcatTTCCATTTCACCGTTGCTC
L45I_fwd	AAATGGAAattAATGAAGCCACCCTGTATACG	L45I_rv	GGCTTCATTaatTTCCATTTCACCGTTGC
E47P_fwd	GAACTGAATccaGCCACCCTGTATACGATTTTTGA	E47P_rv	CAGGGTGGCtggATTCAGTTCCATTTCACC
A48G_fwd	TGAATGAAggcACCCTGTATACGA	A48G_rv	TACAGGGTgccTTCATTCAGTTCCATTTCAC
T52P_fwd	ACCCTGTATccgATTTTTGATCGTCTGG	T52P_rv	CAAAAATcggATACAGGGTGGCTTC
I53V_fwd	CTGTATACGgtgTTTGATCGTCTGGAACAGG	I53V_rv	ACGATCAAAcacCGTATACAGGGTGGCTTC
I53S_fwd	CTGTATACGtctTTTGATCGTCTGGAACAGG	I53S_rv	CGATCAAAagaCGTATACAGGGTGGCTTCA
I53A_fwd	CTGTATACGgcgTTTGATCGTCTGGAACAGGACG	I53A_rv	CGATCAAAcagCGTATACAGGGTGGCTTC

I53L_fwd	CTGTATACGctgTTTGATCGTCTGGAACAGGAC	I53L_rv	ACGATCAAAcagCGTATACAGGGTGGCTTC
F54L_fwd	TATACGATTctgGATCGTCTGGAACAGGACGG	F54L_rv	CCAGACGATCcagAATCGTATACAGGGTGG
K55R_fwd	ATACGATTTTTcgtCGTCTGGAACAGGACGG	K55R_rv	CCAGACGacgAAAAATCGTATACAGGGTGG
K55S_fwd	ATACGATTTTTtctCGTCTGGAACAGGACGG	K55S_rv	TTCCAGACGagaAAAAATCGTATACAGGGTGG
E58Q_fwd	GATCGTCTGcagCAGGACGGCATTATCAG	E58Q_rv	GCCGTCCTGctgCAGACGATCAAAAATC
I62Y_fwd	CAGGACGGCtatATCAGCTCTTACTGGG	I62Y_rv	GAGCTGATataGCCGTCCTGTTCCAG
I62L_fwd	CAGGACGGCctgATCAGCTCTTACTGGG	I62L_rv	GAGCTGATgacGCCGTCCTGTTCCAG
I62W_fwd	AGGACGGCtggATCAGCTCTTACTGGGG	I62W_rv	AGAGCTGATccaGCCGTCCTGTTCCAGAC
S64T_fwd	GGCATTATCaccTCTTACTGGGGTGATGA	S64T_rv	CAGTAAGAggtGATAATGCCGTCC
S64E_fwd	CGGCATTATCgaaTCTTACTGGGGTGATGAAAGTC	S64E_rv	CCCAGTAAGAttcGATAATGCCGTCCTGTTCC
S65T_fwd	ATTATCAGCaccTACTGGGGTGATGAAAGTCAAG	S65T_rv	CCCCAGTAggtGCTGATAATGCCGTC
S65G_fwd	ATTATCAGCggcTACTGGGGTGATGAAAGTCAAGG	S65G_rv	CCCCAGTAgccGCTGATAATGCCGTCCTG
Y66R_fwd	ATCAGCTCTcgtTGGGGTGATGAAAGTCAAGG	Y66R_rv	CATCACCCCAacgAGAGCTGATAATGCCGT
R75P_fwd	AAGGCGGTccgCGCAAATATTACCGTCTGA	R75P_rv	AATATTTGCGcggACCGCCTTGACTTTCATC
R80S_fwd	CAAATATTACagcCTGACCGAAATCGGCCATG	R80S_rv	TTTCGGTCAGgctGTAATATTTGCGACGACC
T82S_fwd	CGTCTGagcGAAATCGGCC	T82S_rv	CCGATTTCgctCAGACGGTAATATTTGCG
I84A_fwd	GACCGAAgcgGGCCATGAAAACATGCGC	I84A_rv	TCATGGCCcgcTTCGGTCAGACGGTAATATTTGC
H86E_fwd	GAAATCGGCgaaGAAAACATGCGCCTG	H86E_rv	GCATGTTTTCttcGCCGATTTCGGTCAGACG
H86K_fwd	GAAATCGGCaaaGAAAACATGCGCCTG	H86K_rv	GCATGTTTTCtttGCCGATTTCGGTCAGACG
H86R_fwd	AATCGGCcgtGAAAACATGCG	H86R_rv	TGTTTTCacgGCCGATTTCGGTCAG

N88E_fwd	GCCATGAAgaaATGCGCCTGCGG	N88E_rv	AGGCGCATttcTTCATGGCCGATTTCG
N88Q_fwd	GCCATGAAcagATGCGCCTGCGG	N88Q_rv	AGGCGCATctgTTCATGGCCGATTTCG
R90Q_fwd	AAACATGcagCTGCGGCATGAATCCT	R90Q_rv	TGCCGCAGctgCATGTTTTCATGGC
R90A_fwd	AAACATGgccCTGCGGCATGAATCCTG	R90A_rv	TGCCGCAGggcCATGTTTTCATGGCC
E94A_fwd	CGGCATgcaTCCTGGAGTCGT	E94A_rv	TCCAGGAtgcATGCCGCAGG
E94Q_fwd	TGCGGCATcagTCCTGGAGTCGTGTG	E94Q_rv	CTCCAGGActgATGCCGCAGGCG
S95E_fwd	GGCATGAAgaaTGGAGTCGTGTGGACAAAATCA	S95E_rv	CGACTCCAttcTTCATGCCGCAGGCG
S95D_fwd	GGCATGAAagcTGGAGTCGTGTGGAC	S95D_rv	GACTCCAgctTTCATGCCGCAGGCG
S97Q_fwd	ATGAATCCTGGcagCGTGTGGACAAAATCATTG	S97Q_rv	GTCCACACGctgCCAGGATTCATGCCG
S97K_fwd	GAATCCTGGaaaCGTGTGGACAAAATCATTG	S97K_rv	GTCCACACGtttCCAGGATTCATGCC
V99L_fwd	CTGGAGTCGTagcGACAAAATCATTGAAAATCTGG	V99L_rv	GATTTTGTCgctACGACTCCAGGATTCATGC
K101R_fwd	GTCGTGTGGACcgtATCATTGAAAATCTGGAAGC	K101R_rv	TTTCAATGATacgGTCCACACGACTCCAGGATTC
I102V_fwd	GTGGACAAAgtgATTGAAAATCTGGAAGCAAAC	I102V_rv	GATTTTCAATcacTTTGTCCACACGACTCCAGG
E107D_fwd	AAATCTGgatGCAAACAAAAAATCTGAAGC	E107D_rv	TTGTTTGCatcCAGATTTTCAATGATTTTGTC
K110G_fwd	GAAGCAAACggcAAATCTGAAGCGATCAAATCTAG	K110G_rv	GCTTCAGATTTgccGTTTGCTTCCAGATTTTCAAT

Table S7: List of DNA primers used to construct the consensus mutant, carrying six point mutations.

Primer Name	Sequence 3’ 5’	Primer Name	Sequence 3’ 5’
Cons_A11G_fwd	TGCGTggtCAAACCAATTAGATCC	Cons_A11G_rv	GGTTTGaccACGCAGCATTTCTTTCG
Cons_A48G_T52P_F54L_fwd	CCCTGTATccgATTctgGATCGTCTGGAACAGGAC	Cons_A48G_T52P_F54L_rv	AATcggATACAGGGTgccTTCATTCAGTTCC
Cons_V28G_fwd	CTATggcTATGGCATTATCAAACAGGTGAAAGAAG	Cons_V28G_rv	ATGCCATAaccATAGTTATCGCCTTGTTTCAGG
Cons_M89L_fwd	AAAACctgCGCCTGCGGC	Cons_M89L_rv	GGCGcagGTTTTCATGGCC

Table S8: List of DNA primers used for combinatorial library preparation.

Primer Name	Sequence 3' 5'	Primer Name	Sequence 3' 5'
I62NDT_fwd	GGACGGCndtATCAGCTCTTACTGG	I62NDT_rv	GAGCTGATahnGCCGTCCTGTTCCA
I62W_fwd_2	AGGACGGCtggATCAGCTCTTACTGGGGT	I62W_rv_2	AGAGCTGATccaGCCGTCCTGTTCCAGAC
N88VHG_fwd	CATGAAvhgATGCGCCTGCGGC	N88VHG_rv	CAGGCGCATcdbTTCATGGCCGATTTCG
N88Q_fwd_2	CCATGAAcagATGCGCCTGCGGC	N88Q_rv_2	CAGGCGCATctgTTCATGGCCGATTTCG

DNA sequence of LmrR_RMH

In red the amber stop codon used for the incorporation of the non-canonical amino acid p-amino phenylalanine is highlighted.

ATGGGTGCCGAAATCCCGAAAGAAATGCTGCGTGCTCAAACCAATTAGATCCTGCTGATGGTCCTGAAACAAG
GCGATAACTATGTGTATGGCATTATCAAACAGGTGAAAGAAGCGAGCAACGGTGAAATGGAACTGAATGAAG
CCACCCTGTATACGATTTTTGATCGTCTGGAACAGGACGGCATTATCAGCTCTTACTGGGGTGATGAAAGTCAA
GGCGGTCGTCGCAAATATTACCGTCTGACCGAAATCGGCCATGAAAACATGCGCCTGCGGCATGAATCCTGGA
GTCGTGTGGACAAAATCATTGAAAATCTGGAAGCAAACAAAAAATCTGAAGCGATCAAATCTAGAGGTGGCA
GCGGTGGCTGGAGCCACCCGCAGTTCGAAAAATAA

Protein sequence of LmrR_RMH

In red the non-canonical amino acid p-amino phenylalanine is highlighted. Underlined the three mutations (A92R_N19M_F93H) that characterise the triple mutant LmrR_RMH are reported.

GAEIPKEMLRAQTNpAFILLMVLKQGDNYVYGIIKQVKEASNGEMELNEATLYTIFDRLEQDGIISSYWGDESQGGR
RKYYRLTEIGHENMRLRHESWSRVDKIIENLEANKKSEAIKSRGGSGGWSHPQFEK

References

1 M. Amiram, A. D. Haimovich, C. Fan, Y.-S. Wang, H.-R. Aerni, I. Ntai, D. W. Moonan, N. J. Ma, A. J. Rovner, S. H. Hong, N. L. Kelleher, A. L. Goodman, M. C. Jewett, D. Söll, J. Rinehart and F. J. Isaacs, *Nat Biotechnol*, 2015, **33**, 1272–1279.

2 M. R. Green and J. Sambrook, *Cold Spring Harb Protoc*, 2020, **2020**, 101196.

3 D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison and H. O. Smith, *Nat Methods*, 2009, **6**, 343–345.

4 L. A. Mitchell, Y. Cai, M. Taylor, A. M. Noronha, J. Chuang, L. Dai and J. D. Boeke, *ACS Synthetic Biology*, 2013, **2**, 473–477.

5 E. Krieger and G. Vriend, *Bioinformatics (Oxford, England)*, 2014, **30**, 2981–2982.

6 D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, N. Forouzesh, T. J. Giese, A. W. Götz, H. Gohlke, S. Izadi, K. Kasavajhala, M. C. Kaymak, E. King, T. Kurtzman, T.-S. Lee, P. Li, J. Liu, T. Luchko, R. Luo, M. Manathunga, M. R. Machado, H. M. Nguyen, K. A. O'Hearn, A. V. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, A. Risheh, S. Schott-Verdugo, A. Shajan, J. Swails, J. Wang, H. Wei, X. Wu, Y. Wu, S. Zhang, S. Zhao, Q. Zhu, T. E. I. Cheatham, D. R. Roe, A. Roitberg, C. Simmerling, D. M. York, M. C. Nagan and K. M. Jr. Merz, *J. Chem. Inf. Model.*, 2023, **63**, 6183–6191.

7 F. Madeira, Y. mi Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn and R. Lopez, *Nucleic Acids Research*, 2019, **47**, W636–W641.

8 R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane and V. S. Pande, *Biophys J*, 2015, **109**, 1528–1532.

9 D. R. Roe and T. E. I. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.

10 H. Nguyen, D. R. Roe, J. Swails and D. A. Case, PYTRAJ: Interactive data analysis for molecular dynamics simulations https://amber-md.github.io/pytraj/latest/index.html 2016.

11 R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, J. Domański, D. L. Dotson, S. Buchoux, I. M. Kenney and O. Beckstein, *Proceedings of the 15th Python in Science Conference*, 2016, 98–105.

12 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *Journal of Computational Chemistry*, 2011, **32**, 2319–2327.

13 S. van der Walt, S. C. Colbert and G. Varoquaux, *Computing in Science and Engineering*, 2011, **13**, 22–30.

14 M. Canyelles-Niño, RCBS.py: Reactivity of Chemical and Biochemical Systems (version 2.0.4) https://github.com/MolBioMedUAB/RCBS.py 2022.

15 The pandas development team, pandas-dev/pandas: Pandas (version 2.2.1) https://github.com/pandas-dev/pandas.

16 J. D. Hunter, *Computing in Science & Engineering*, 2007, **9**, 90–95.

17 M. L. Waskom, *Journal of Open Source Software*, 2021, **6**, 3021.

18 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J Comput Chem*, 2004, **25**, 1605–1612.

19 J. Shao, S. W. Tanner, N. Thompson and T. E. Cheatham, *J. Chem. Theory Comput.*, 2007, **3**, 2312–2334.

20 M. Canyelles-Niño, MolBioMedUAB/EMDA (version 1.0.0a4) https://doi.org/10.5281/zenodo.10813141 2024.

21 M. C. R. Melo, R. C. Bernardi, C. de la Fuente-Nunez and Z. Luthey-Schulten, *The Journal of Chemical Physics*, 2020, **153**, 134104.

22 G. E. Crooks, G. Hon, J.-M. Chandonia and S. E. Brenner, *Genome Res*, 2004, **14**, 1188–1190.

23 T. D. Schneider and R. M. Stephens, *Nucleic Acids Res*, 1990, **18**, 6097–6100.