Supplementary Information: Assessing the capability of *in silico* mutation protocols for predicting the finite temperature conformation of amino acids

Rodrigo Ochoa[1], Miguel A. Soler[2], Alessandro Laio[2,3], Pilar Cossio[1,4]

[1]Biophysics of Tropical Diseases, Max Planck Tandem Group, University of Antioquia, Medellín, Colombia.

[2]International School for Advanced Studies (SISSA), Via Bonomea 265, I-34136 Trieste, Italy.

[3]The Abdus Salam International Centre for Theoretical Physics (ICTP), Strada Costiera 11, 34151 Trieste, Italy.

[4]Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany.

Supplementary Text and Figures

Example of wrong rotamer effects in the Barnase-Barstar system

We provide an example of the effects of wrong rotamer placement for the Barnase-Barstar system. We mutated alanine residue A27 by a lysine in the Barnase chain using different rotamer programs. One program generated a rotamer that was close to the mutated crystal while the other did not (Figure S1A). We performed three MD simulations of the system, two starting from the predicted rotamers and one from the protein crystal containing the mutation. In Figure S1B, we show the number of contacts as a function of the simulation time. We notice that the predicted rotamer that was in an opposite direction with respect to the crystal, lost a considerable amount of native contacts within the interaction.

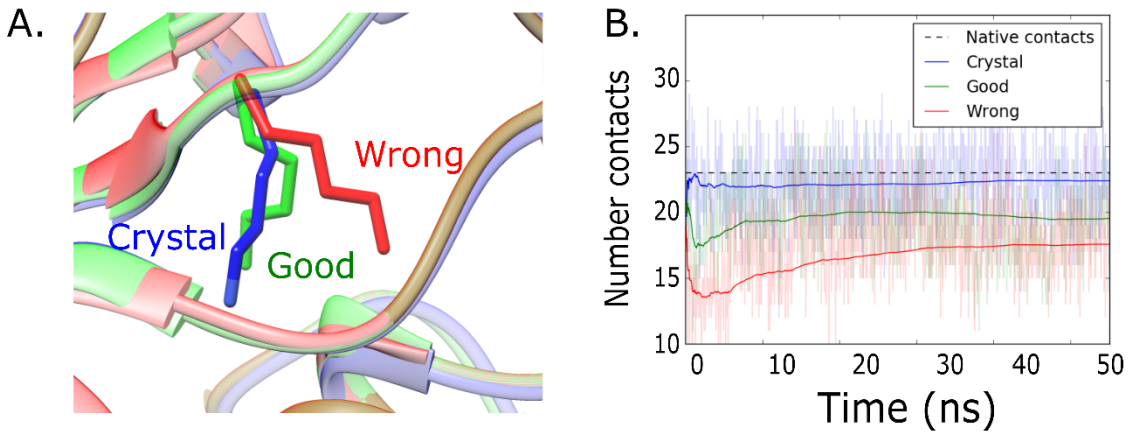

Figure S1. (A) Structural view of the crystal rotamer (blue) and the predicted rotamers (green and red). (B) Number of contacts for the mutated Barnase-Barstar system (AA27K) as a function of the simulation time for the original crystal (blue), the complex with a predicted mutation close to the crystal rotamer (green), and the complex with other predicted mutation in an opposite direction to the crystal rotamer (red).

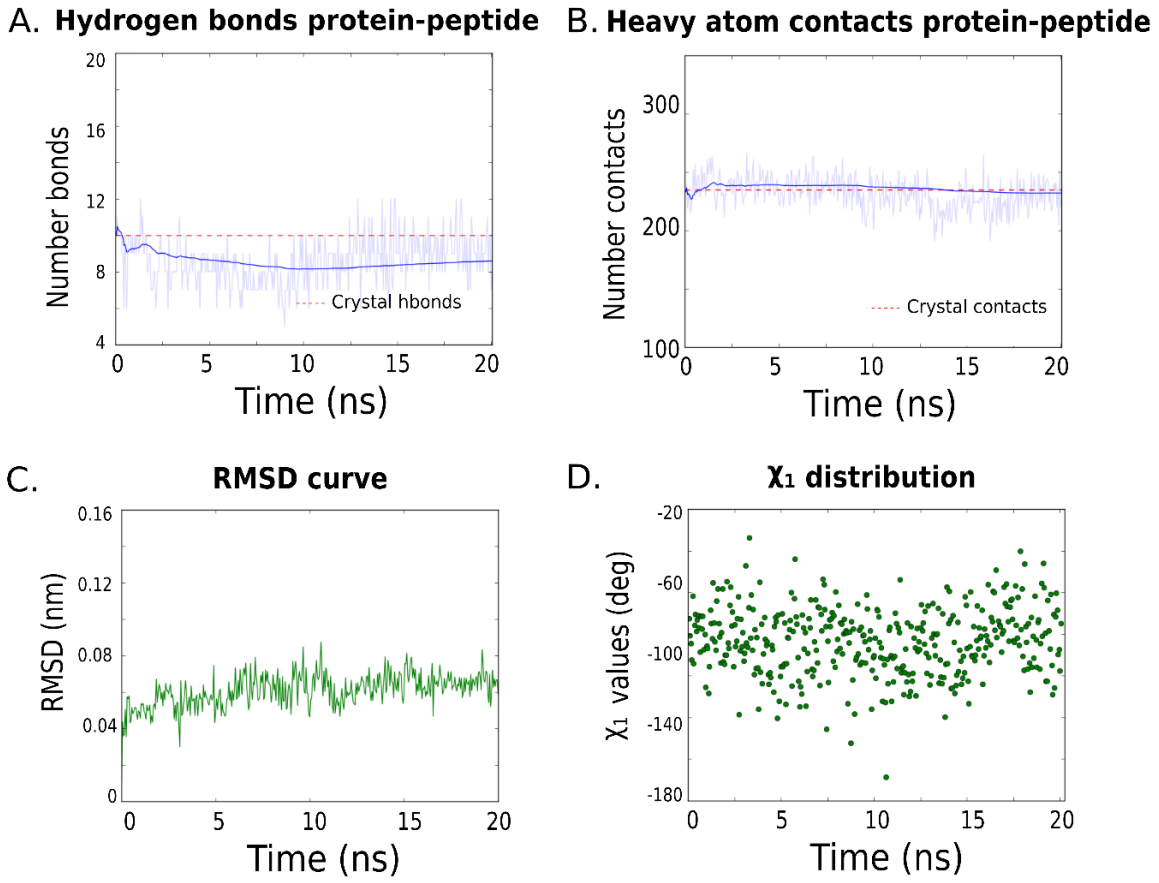

Figure S2. Example of the convergence from the molecular dynamics (MD) simulations. Observables that monitor the convergence of the OppA system complexed with the K-F-K tripeptide (PDB:1B40). (A) Number of hydrogen bonds (blue line) between the peptide and the protein during the trajectory. The number of hydrogen bonds for the starting crystal is shown as a dashed red line. (B) Number of heavy-atom contacts between the peptide and the protein. The number of contacts for the starting crystal is shown as a dashed red line. (C) All-atom RMSD of the peptide during the 20 nanoseconds. (D) Distribution of the χ_1 dihedral angle for the mutated amino acid.

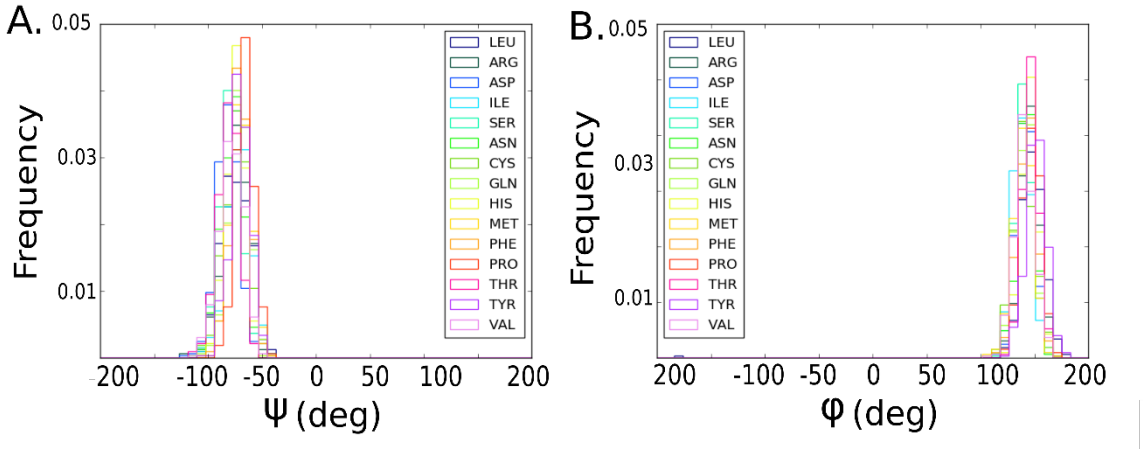

Figure S3. Distribution of ψ (A) and φ (B) backbone dihedral angles for the mutated amino acids studied in the OppA system, from the molecular dynamics simulations.

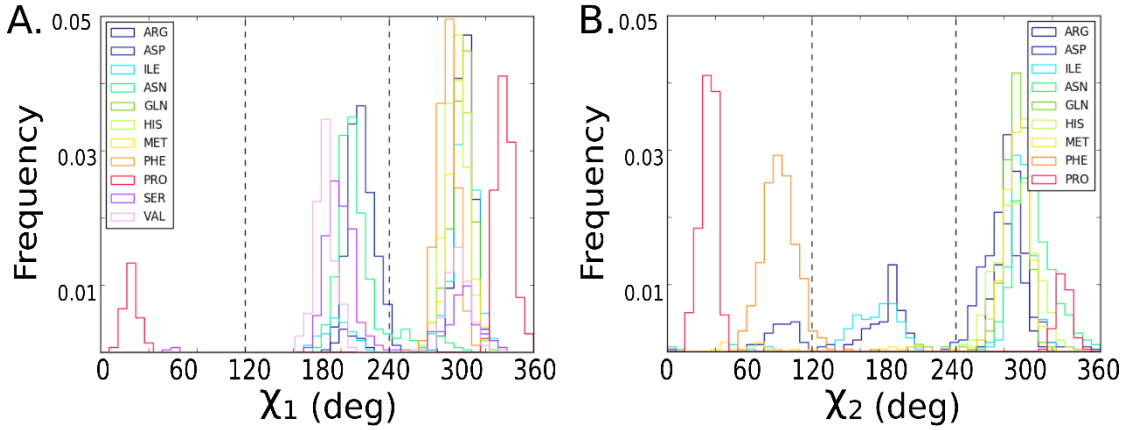

Figure S4. Distribution of χ_1 (A) and χ_2 (B) dihedral angles of the mutated amino acids studied in the OppA system, from the molecular dynamics simulations. The histograms were categorized in three main regions containing the *gauche*(-) (0° to 120°), *trans* (120° to 240°) and *gauche*(+) (240° to 360°) side chain conformations, as indicated by the dashed blue lines.

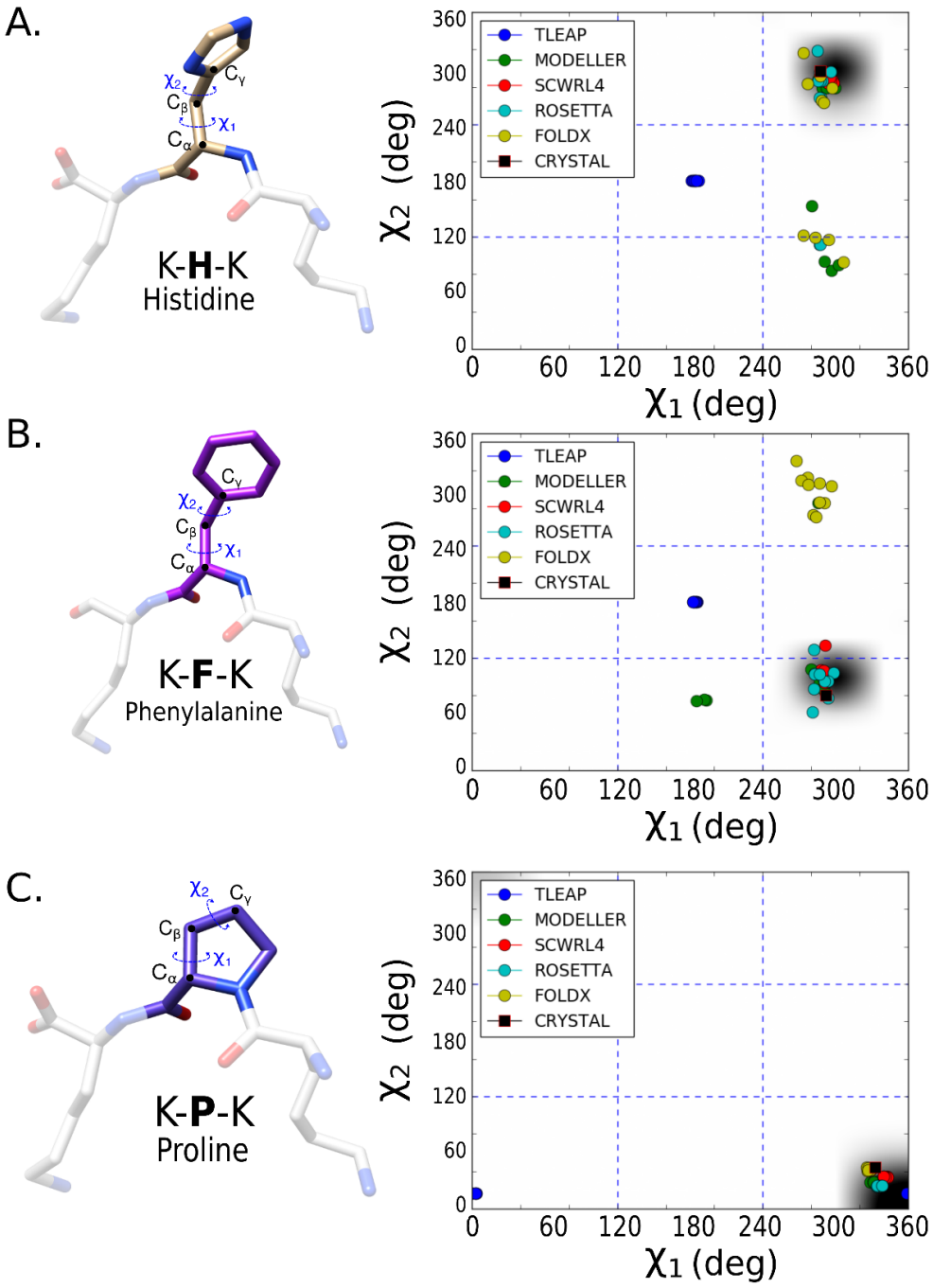

Figure S5. 2D Histogram of χ_1 vs χ_2 dihedrals for three mutated amino acids; histidine (A), phenylalanine (B) and proline (C), on the tripeptide. Each mutation protocol prediction is represented by a circle (see Methods), dihedral values from the crystal structure by a black square, and the distribution from the molecular dynamics is shown as shaded grey. The

main conformational groups are split in 9 regions based on the χ_1 and χ_2 possible combinations.

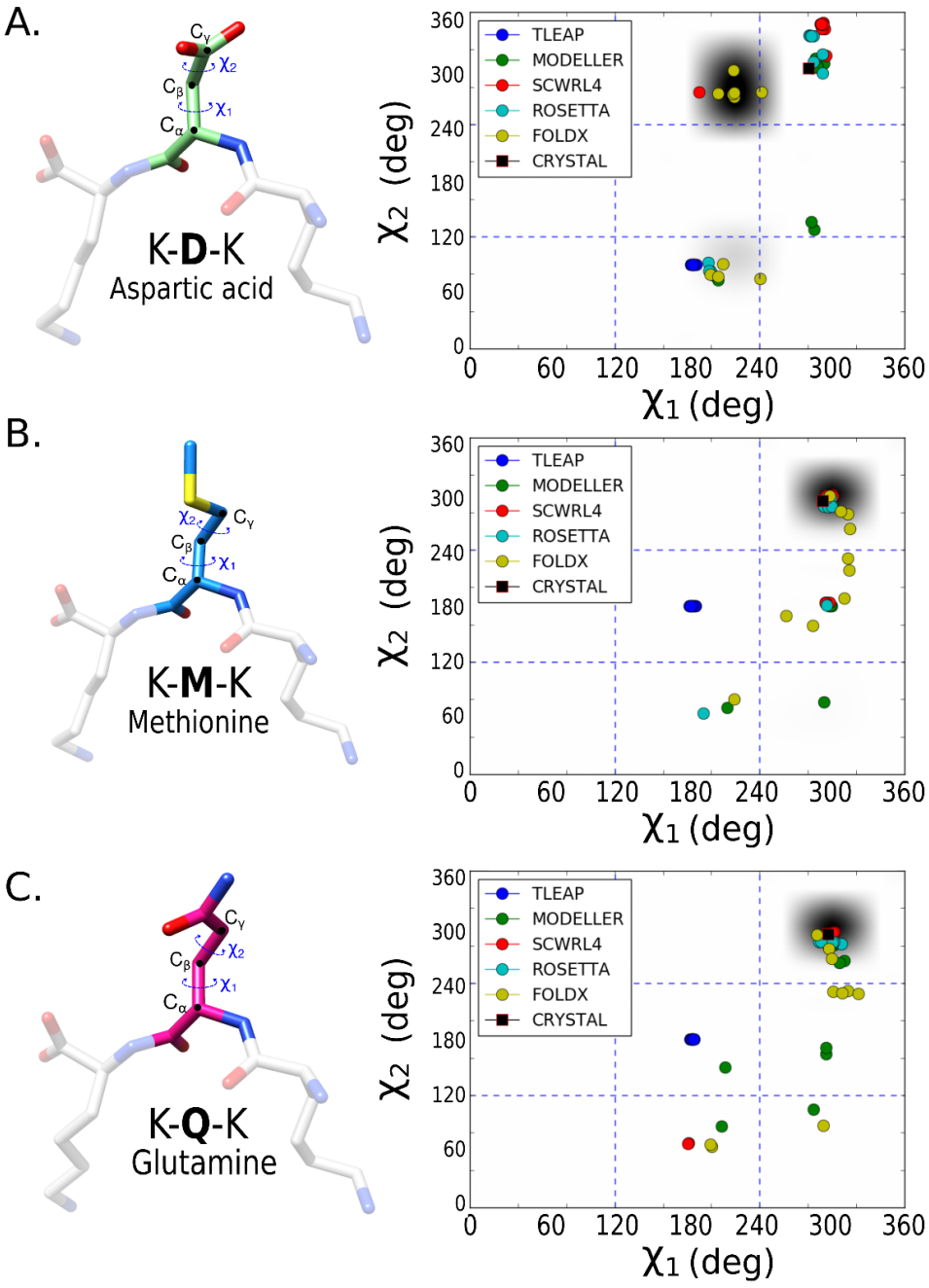

Figure S6. 2D Histogram of χ_1 vs χ_2 dihedrals for three variable amino acids: aspartic acid (A), methionine (B) and glutamine (C), on the tripeptide. Each mutation protocol prediction is represented by a circle (see Methods), dihedral values from the crystal structure by a

black square, and the distribution from the molecular dynamics is shown as shaded grey. The main conformational groups are split in 9 regions based on the χ_1 and χ_2 possible combinations.

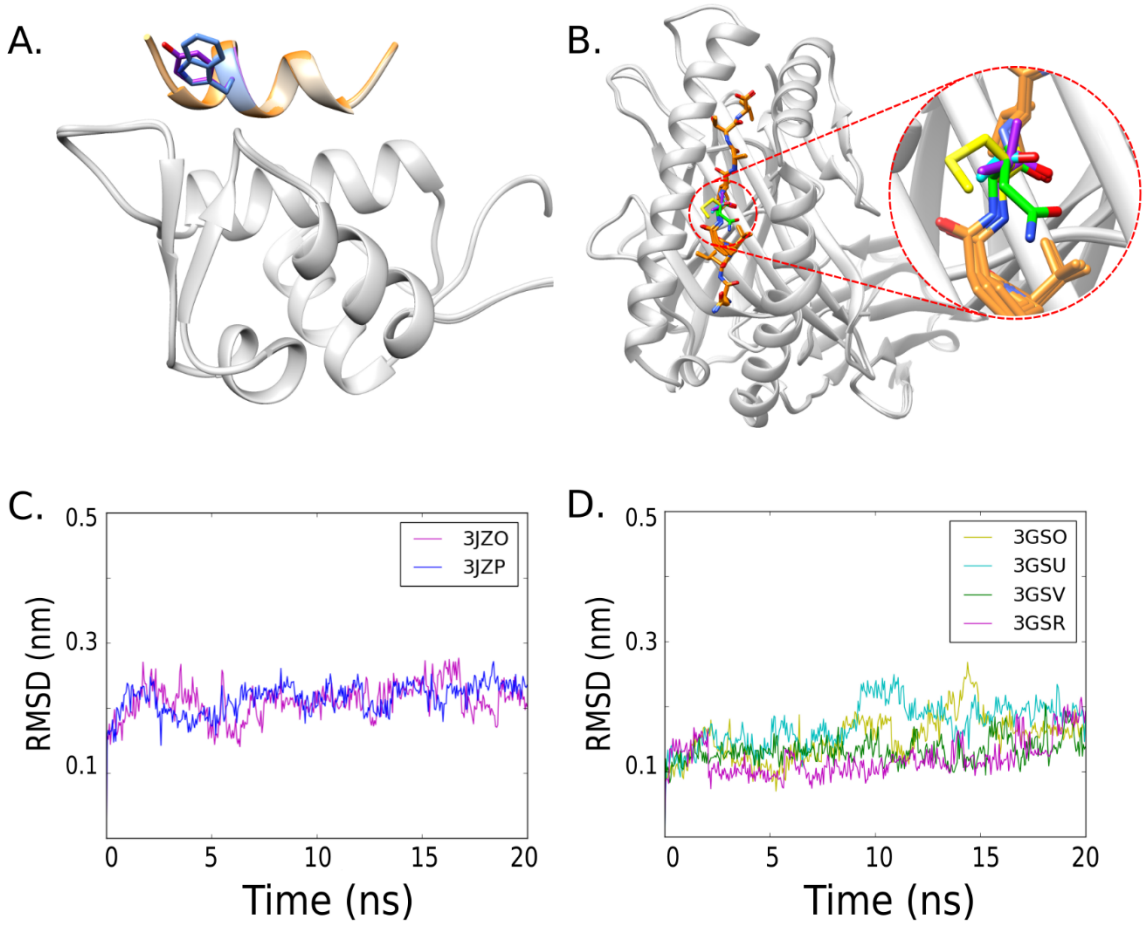

Figure S7. Structural alignment of the two additional protein-peptide complexes used to assess the predictability of the mutation protocols. (A) MDM4 protein complexed with a 11-mer peptide (orange), mutated in the position Tyr6 (purple) by Trp6 (blue). (B) HLA class I protein in complex with a 9-mer peptide, with crystal structures containing four single mutants at the 5th position: Met (yellow), Gln (green), Val (purple) and Thr (cyan). (C) All-atom RMSD of the peptide as a function of the simulation time, for the two variations bound to MDM4. (D) All-atom RMSD of the peptide for the four variations of the peptides bound to the HLA class I protein.

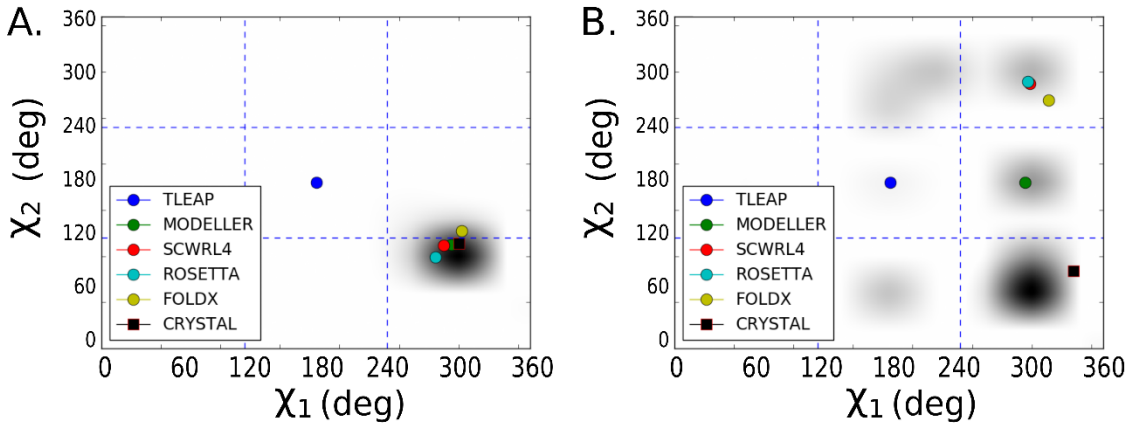

Figure S8. 2D histogram of χ_1 vs χ_2 dihedral angles for two protein-peptide systems from the molecular dynamics simulations and mutation protocol predictions: (A) the MDM4 complex with mutation YP6W and (B) the HLA class I complex with the mutation TP5Q. Each mutation protocol prediction is represented by a circle (see Methods), dihedral values from the crystal structure by a black square, and the main conformation groups are split in 9 regions based on the χ_1 and χ_2 possible combinations.

Supplementary Tables

Table S1. Average of the success rate for the χ_1 and χ_2 dihedral angle prediction per mutation protocol and amino acid using a 30% threshold.

AA	MODELLER	SCWRL4	TLEAP	ROSETTA	FOLDX
ARG	0	40	0	50	30
ASN	0	0	0	0	30
ASP	0	0	0	0	50
GLN	20	80	0	70	0
HIS	30	10	0	60	10
ILE	50	10	0	50	50
MET	50	60	0	80	20
PHE	60	90	0	70	0
PRO	0	100	0	100	90
Average	**23.3**	**43.3**	**0.0**	**53.3**	**31.1**

Table S2. Success rate average for the χ_1 and χ_2 dihedral angle prediction per mutation protocol generated from amino acid conformation from the last MD frame (instead of the crystals) and using a 5% threshold.

AA	MODELLER	SCWRL4	TLEAP	ROSETTA	FOLDX
ARG	40	50	30	50	60
ASN	40	90	0	80	40
ASP	10	20	0	10	20
GLN	50	30	0	60	40
HIS	40	80	0	80	70
ILE	100	100	60	90	60
MET	20	40	0	50	40
PHE	70	80	0	90	30
PRO	100	100	100	100	100
Average	**52.2**	**65.6**	**21.1**	**67.8**	**51.1**

Table S3. Success rate average for the χ_1 and χ_2 dihedral angle prediction per mutation protocol from the perspective of the initial amino subjected to mutation (using a 5% threshold).

AA	MODELLER	SCWRL4	TLEAP	ROSETTA	FOLDX
ARG	50	60	10	50	70
ASN	40	60	10	60	40
ASP	40	80	10	80	20
GLN	40	60	10	60	50
HIS	50	50	10	50	30
ILE	30	50	10	50	40
MET	50	60	10	60	40
PHE	40	60	10	70	40
PRO	20	60	0	70	30
SER	30	70	10	40	40
VAL	20	70	10	80	50
Average	**37.3**	**61.8**	**9.1**	**60.9**	**40.9**

Table S4. Average contact conservation using a 3.5Å threshold-distance between the predicted mutation and structures from the molecular dynamics simulations with similar χ_1 and χ_2 dihedral angles.

AA	MODELLER	SCWRL4	TLEAP	ROSETTA	FOLDX
ARG	0	0.34 ± 0.02	0	0.44 ± 0.02	0.18 ± 0.03
ASN	0.28 ± 0.04	0	0	0.06 ± 0.03	0.03 ± 0.01
ASP	0.11 ± 0.03	0	0	0.14 ± 0.04	0.63 ± 0.04
PHE	0.40 ± 0.05	0.59 ± 0.03	0	0.51 ± 0.04	0
HIS	0.27 ± 0.04	0.59 ± 0.03	0	0.38 ± 0.03	0.29 ± 0.045
ILE	0.50 ± 0.04	0.79 ± 0.03	0	0.54 ± 0.06	0.58 ± 0.06
MET	0.35 ± 0.05	0.39 ± 0.05	0	0.56 ± 0.05	0.23 ± 0.05
PRO	0.18 ± 0.05	0.72 ± 0.03	0.59 ± 0.02	0.75 ± 0.03	0.83 ± 0.03
GLN	0.25 ± 0.04	0.54 ± 0.05	0	0.67 ± 0.03	0.14 ± 0.04
SER	0.06 ± 0.03	0.24 ± 0.04	0.65 ± 0.02	0.06 ± 0.03	0.46 ± 0.05
VAL	0.73 ± 0.02	0.72 ± 0.03	0	0.74 ± 0.03	0.78 ± 0.03
Average	**0.28 ± 0.06**	**0.44 ± 0.08**	**0.11 ± 0.07**	**0.44 ± 0.08**	**0.38 ± 0.09**

Table S5. Average contact conservation using a 4Å threshold-distance between the predicted mutation and structures from MD with similar χ_1 and χ_2 dihedral angles.

AA	MODELLER	SCWRL4	TLEAP	ROSETTA	FOLDX
ARG	0	0.46 ± 0.02	0	0.55 ± 0.02	0.24 ± 0.04
ASN	0.40 ± 0.06	0	0	0.08 ± 0.03	0.07 ± 0.03
ASP	0.14 ± 0.04	0	0	0.14 ± 0.04	0.73 ± 0.04
PHE	0.43 ± 0.05	0.61 ± 0.03	0	0.58 ± 0.04	0
HIS	0.40 ± 0.05	0.67 ± 0.01	0	0.54 ± 0.04	0.34 ± 0.05
ILE	0.65 ± 0.05	0.85 ± 0.01	0	0.56 ± 0.06	0.64 ± 0.06
MET	0.38 ± 0.05	0.44 ± 0.05	0	0.57 ± 0.04	0.24 ± 0.05
PRO	0.16 ± 0.05	0.81 ± 0.01	0.74 ± 0.01	0.89 ± 0.01	0.90 ± 0.01
GLN	0.33 ± 0.05	0.60 ± 0.05	0	0.69 ± 0.02	0.16 ± 0.05
SER	0.08 ± 0.03	0.27 ± 0.05	0.81 ± 0.01	0.08 ± 0.03	0.55 ± 0.05
VAL	0.88 ± 0.01	0.80 ± 0.01	0	0.88 ± 0.01	0.85 ± 0.01
Average	**0.35 ± 0.08**	**0.50 ± 0.09**	**0.14 ± 0.09**	**0.51 ± 0.09**	**0.43 ± 0.10**