

Supporting Information.

Exhaustive Exploration of the Conformational Landscape of Small Cyclic Peptides using a Robotics Approach

Maud Jusot^{1,2}, Dirk Stratmann¹, Marc Vaisset², Jacques Chomilier¹, Juan Cortés²

Corresponding authors :

— jacques.chomilier@upmc.fr

— juan.cortes@laas.fr

1. Sorbonne Université, MNHN, CNRS, IRD, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, IMPMC, 75005 Paris, France

2. LAAS-CNRS, Université de Toulouse, CNRS, 31400 Toulouse, France

1 Convergence of Replica Exchange Molecular Dynamics (REMD) simulations

The convergence of Replica Exchange Molecular Dynamics (REMD) simulations was verified for the three peptides (RGDFV, RGDfV, RGDfV') by cutting the trajectory of the lower temperature replica into four equal parts. The Ramachandran diagram was plotted for each part and each residue (see Figures S1, S2, S3 respectively). The explored area of the Ramachandran maps were compared for each part of the trajectory to ensure that the simulation time was long enough. They are reported in Table S1. The percentages of acceptance of exchanges between replica were also computed (see Table S2).

TABLE S1 – Convergence of the REMD trajectories. Level of coverage of the Ramachandran maps explored by the simulation for the three peptides of this study. The range corresponds to the four equally long trajectory parts.

sequence	1st residue	2nd residue	3d residue	4th residue	5th residue
RGDFV	42%-45%	48%-50%	31%-32%	34%-35%	29%-30%
RGDfV	37%-39%	30%-31%	15%-16%	13%-14%	17%-19%
RGDfV'	38%-40%	35%-36%	18%-20%	11%-12%	14%-16%

TABLE S2 – Average probabilities of accepting exchanges between neighboring replica.

Sequence	1st-2nd	2nd-3d	3d-4th	4th-5th	5th-6th	6th-7th	7th-8th
RGDFV	78%	75%	73%	72%	69%	67%	64%
RGDfV	78%	75%	74%	72%	69%	67%	64%
RGDfV'	78%	75%	73%	72%	68%	67%	64%

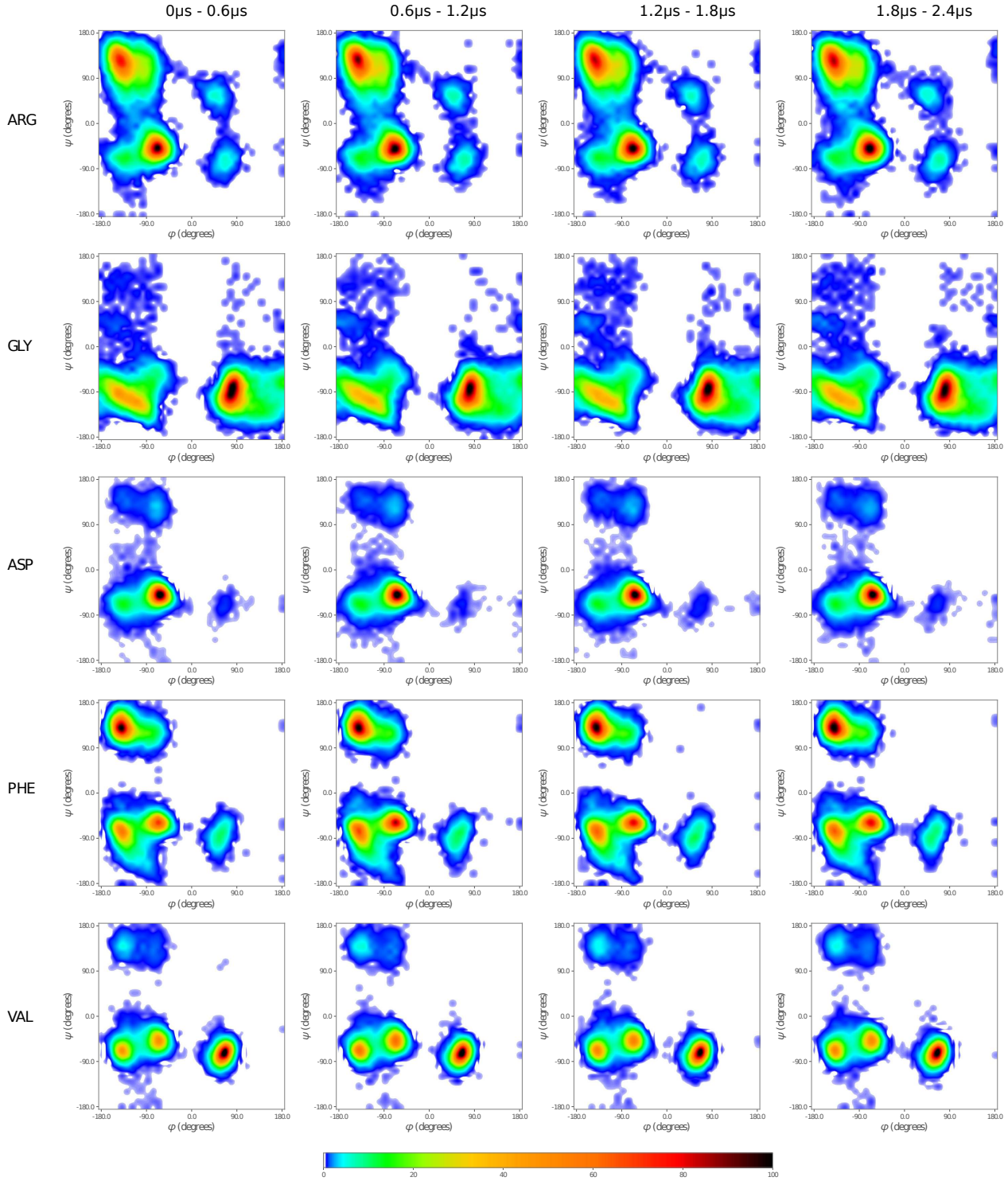


FIGURE S1 – REMD trajectories convergence represented on Ramachandran diagrams for RGDFV peptide. Each column corresponds to a part of the trajectory for the lowest temperature replica, and each line corresponds to a residue. The color code corresponds to normalized frequencies ($nfreq_i = 100 - 100 \cdot \frac{max_{freq} - freq_i}{max_{freq}}$) of conformations found during the simulations. The diagrams have been created using the matplotlib python library¹.

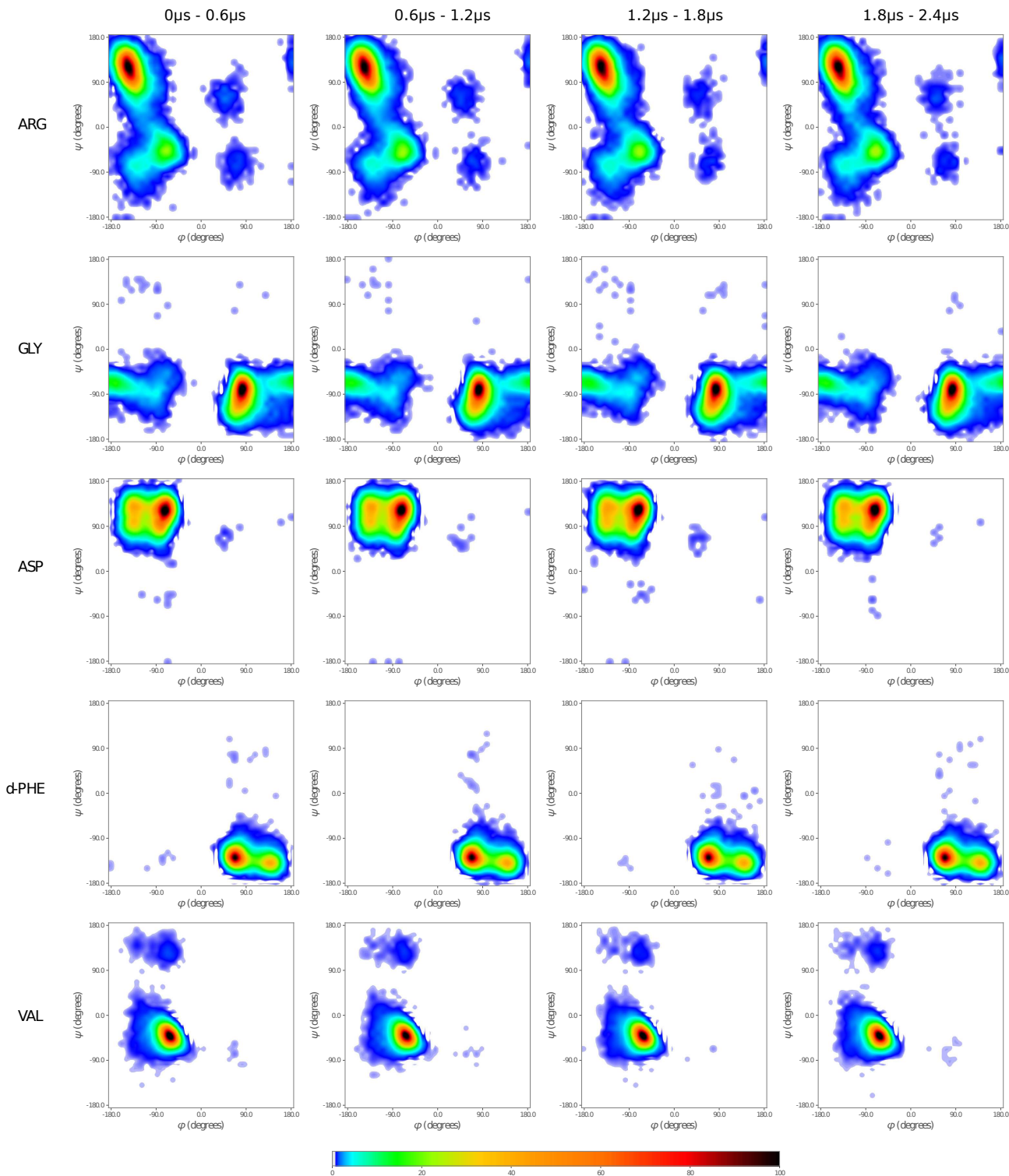


FIGURE S2 – REMD trajectories convergence represented on Ramachandran diagram for RGDfV peptide. Each column corresponds to a part of the trajectory for the lowest temperature replica, and each line corresponds to a residue. The color code corresponds to normalized frequencies ($nfreq_i = 100 - 100 \cdot \frac{max_freq - freq_i}{max_freq}$) of conformations found during the simulations. The diagrams have been created using the matplotlib python library¹.

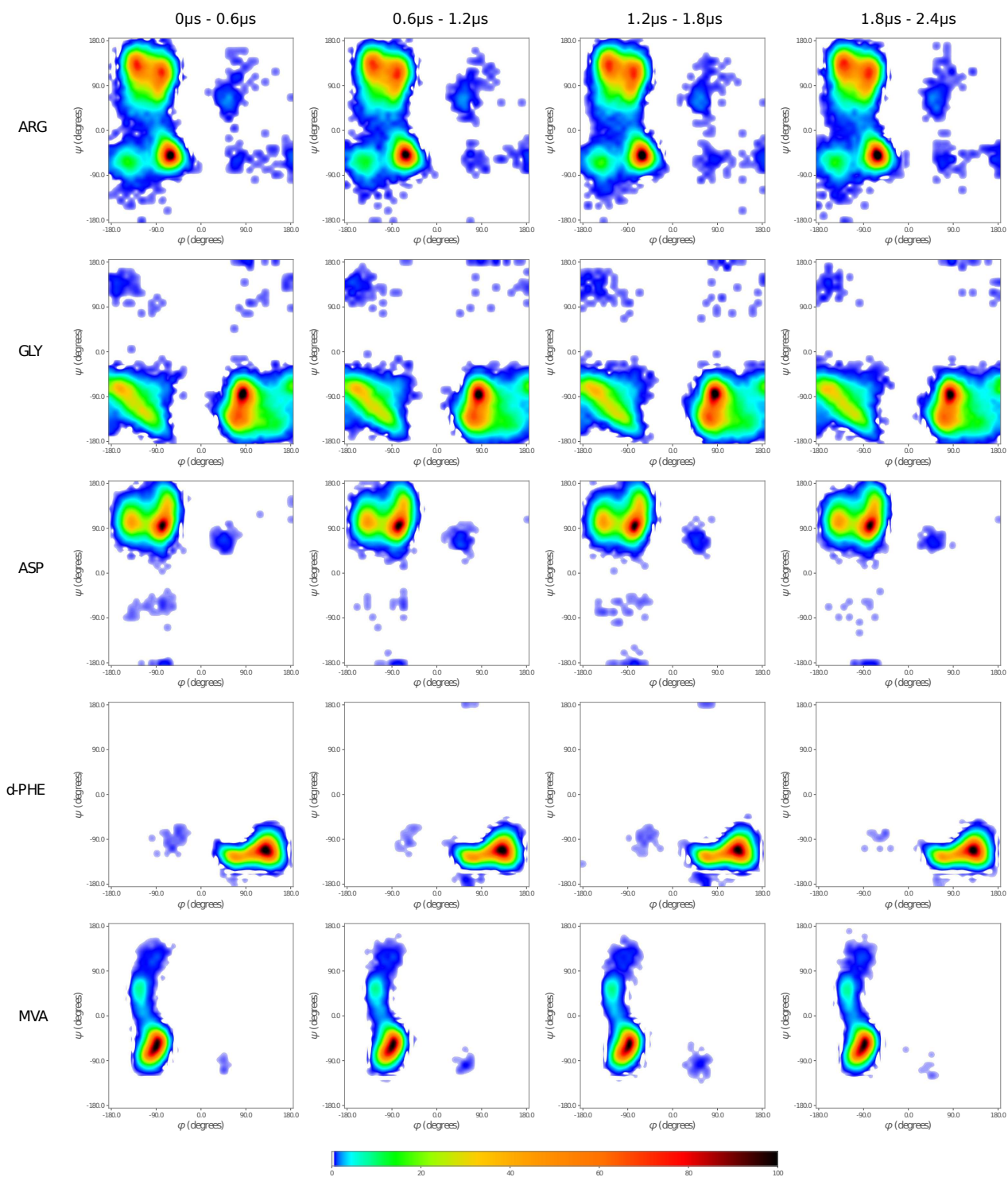


FIGURE S3 – REMD trajectories convergence represented on Ramachandran diagram for RGDfV' peptide (Cilengitide - PDB code : 1L5G). Each column corresponds to a part of the trajectory for the lowest temperature replica, and each line corresponds to a residue. The color code corresponds to normalized frequencies ($nfreq_i = 100 - 100 \cdot \frac{max_{freq} - freq_i}{max_{freq}}$) of conformations found during the simulations. The diagrams have been created using the matplotlib python library¹.

2 Inverse Kinematics (IK)

The number of solution(s) found for each successful call to the IK solver is reported for each peptide in Table S3. In most cases, several solutions are found with an occurrence maximum at two solutions.

TABLE S3 – Number of times the IK solver retrieve n solution(s) (with n from one to ten) for the three peptides of this study.

IK solutions	1	2	3	4	5	6	7	8	9	10
RGDFV - BH/MC	11,174	176,912	3,368	68,073	377	14,894	183	3,102	0	0
RGDFV - SCWRL4	10,990	176,738	3,095	68,045	318	14,987	165	3,042	0	1
RGDfV	10,118	170,655	3,389	76,210	786	20,462	329	4,402	0	0
RGDfV'	20,699	123,378	5,542	45,749	1,065	8,018	132	2,438	0	0

3 Choice of the exhaustively sampled residues

To verify that the choice of the two exhaustively sampled residues does not affect the results of EGSCyP, five explorations were performed selecting all possible five different pairs of consecutive residues for the RGDfV peptide (the other three being treated by the IK solver). Ramachandran diagrams of these explorations are represented in Figure S4. The associated coverage percentages are provided in Table S4.

TABLE S4 – Coverage percentage of the Ramachandran map for each residue of the RGDfV peptide with the sliding window exploration realised with EGSCyP.

sequence	ARG	GLY	ASP	d-PHE	VAL
RGDfV - ARG & VAL sampled	97%	90%	86%	93%	98%
GDFvR - ARG & GLY sampled	97%	98%	94%	90%	94%
DfVRG - ASP & GLY sampled	92%	98%	97%	95%	89%
fVRGD - PHE & ASP sampled	85%	90%	97%	98%	95%
VRGDf - PHE & VAL sampled	90%	85%	90%	98%	98%

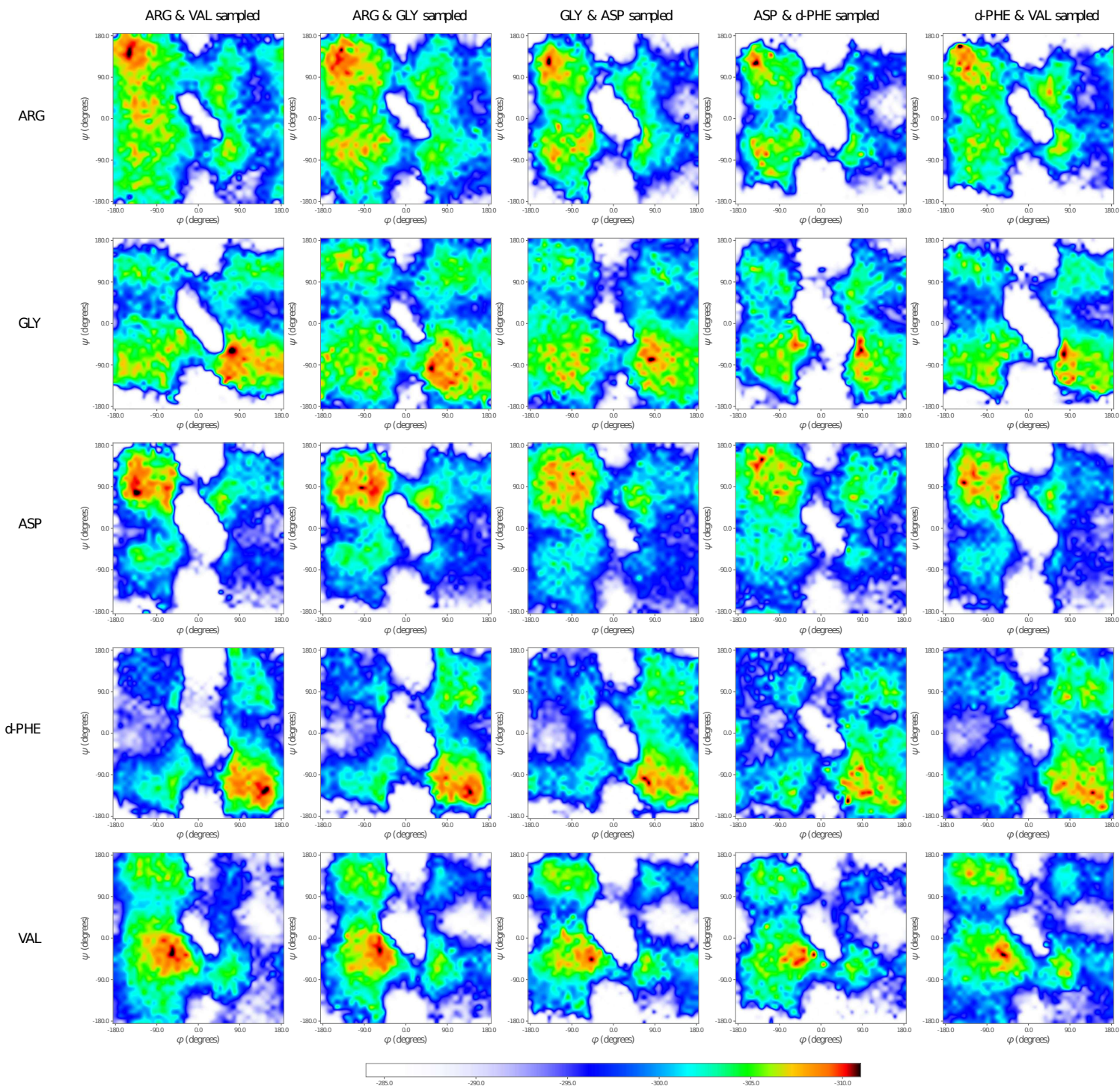


FIGURE S4 – Ramachandran diagrams produced by EGSCyP considering a different selection for the two exhaustively sampled residues on RGDfV peptide. The diagrams were created using the matplotlib python library¹. The energy landscape of the peptide was projected as a function of the dihedral angles ϕ and ψ for each residue (on each line). The color code corresponds to the minimal potential energy in kcal/mol of the whole peptide associated to this corresponding ϕ , ψ pair. Energy minima are represented in black/red while areas with no conformation or high energy are in white/blue.

4 Comparison EGSCyP - REMD for RGDFV and RGDfV peptides

The comparison of the energy landscapes obtained with EGSCyP and REMD was made with Ramachandran diagrams plotted for each peptide and method. The diagrams are presented in Figure S5 for RGDFV and RGDfV peptides. The associated coverage percentages are presented in Table S5, with also the coverage for RGDfV' (associated Ramachandran in Figure 3).

TABLE S5 – Coverage percentage of the Ramachandran diagrams for each residue of (a) RGDFV, (b) RGDfV and (c) RGDfV' peptides, for the EGSCyP method and the REMD simulations.

(a) RGDFV

Method	ARG	GLY	ASP	PHE	VAL
EGSCyP	97%	89%	88%	93%	98%
REMD	53%	62%	41%	41%	31%

(b) RGDfV

Method	ARG	GLY	ASP	d-PHE	VAL
EGSCyP	97%	90%	86%	93%	98%
REMD	46%	38%	20%	19%	24%

(c) RGDfV'

Method	ARG	GLY	ASP	d-PHE	N-Me-VAL
EGSCyP	97%	91%	88%	89%	96%
REMD	49%	45%	25%	14%	18%

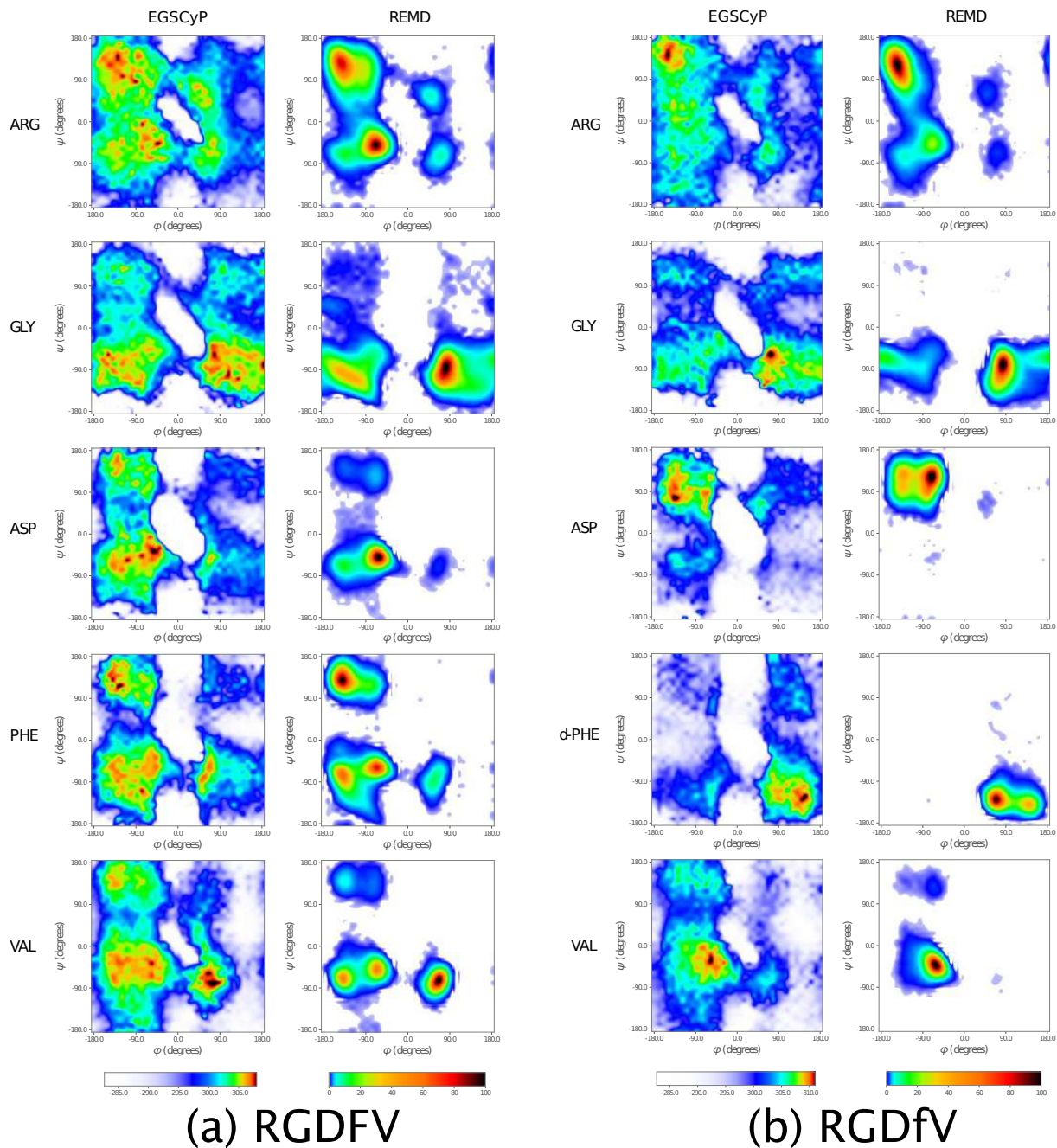


FIGURE S5 – Ramachandran diagrams for each residue of (a) RGDFV and (b) RGdfV peptides. The diagrams were created using the matplotlib python library¹. For EGSCyP (first column of each peptide), the energy landscape was projected as a function of the dihedral angles ϕ and ψ for each residue (on each line). The color code corresponds to the minimal potential energy of the whole peptide associated to this combination of ϕ , ψ . Energy minima in kcal/mol are represented in black/red while areas with no conformation or high energy are in white/blue. For the REMD (second column of each peptide), the normalized frequency ($nfreq_i = 100 - 100 \cdot \frac{maxfreq - freq_i}{maxfreq}$) of the conformations found during the simulations at the lowest temperature replica was projected as a function of ϕ and ψ , with the maximal frequency in black and null frequency in white.

5 Comparison of minima for Cilengitide with EGSCyP and REMD

Figure S7 and Figure S6 show the diversity of the conformations obtained with EGSCyP for Cilengitide. Figure S7 compares the clusters from EGSCyP and REMD simulations. Figure S6 shows the structural representation of the 20 first minima produced by EGSCyP. The pentagonal form of the cycle is maintained in all the different minima (as seen in the top view). However, the residues adopt a wide range of different conformations. The values of the ϕ - ψ angles of the two main minima obtained with both methods (REMD simulations and EGSCyP) are presented in Table S6, showing the correspondence.

TABLE S6 – Values of the ϕ - ψ angles for each minimum.

minimum	phi1	psi1	phi2	psi2	phi3	psi3	phi4	psi4	phi5	psi5
$\text{min}_1^{\text{EGSCyP}}$	-60	-52	-141	-95	-76	75	135	-112	-96	-57
$\text{min}_1^{\text{REMD}}$	-54	-37	170	-71	-61	118	66	-123	-90	-40
$\text{min}_2^{\text{EGSCyP}}$	-125	121	78	-140	-75	86	141	-125	-85	-40
$\text{min}_2^{\text{REMD}}$	-125	122	79	-70	-142	99	120	-105	-97	-38

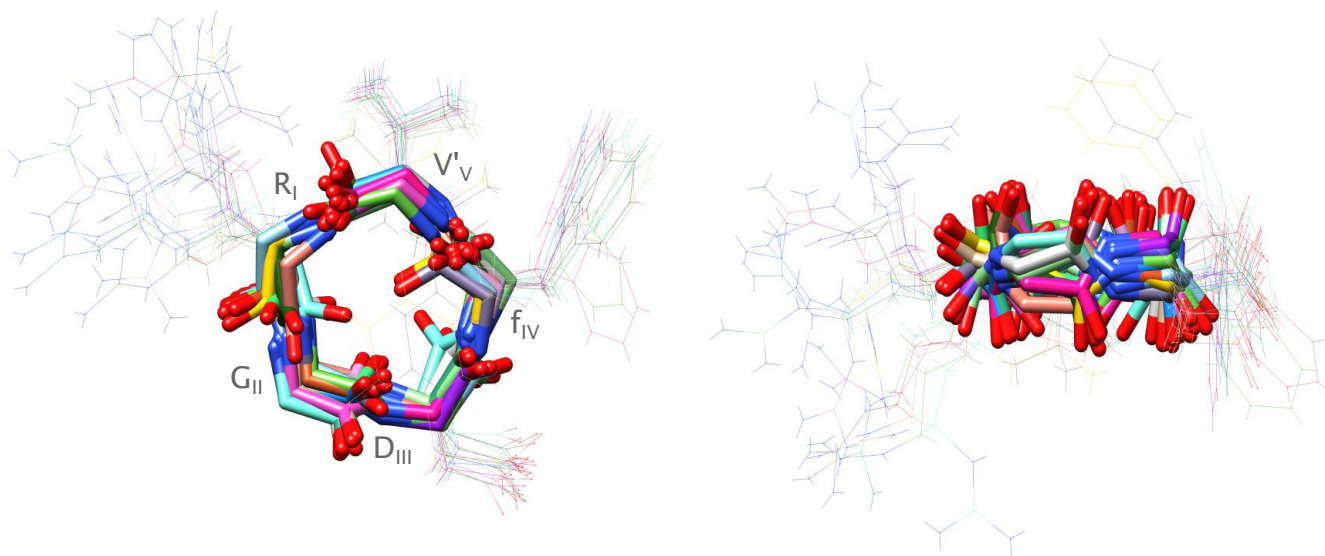


FIGURE S6 – Superimposition of the 20 first minima of Cilengitide from EGSCyP with a top view on the left and a view from the peptide plane on the right. Each color represents a different minimum. For clarity, the side chains are presented in lines.

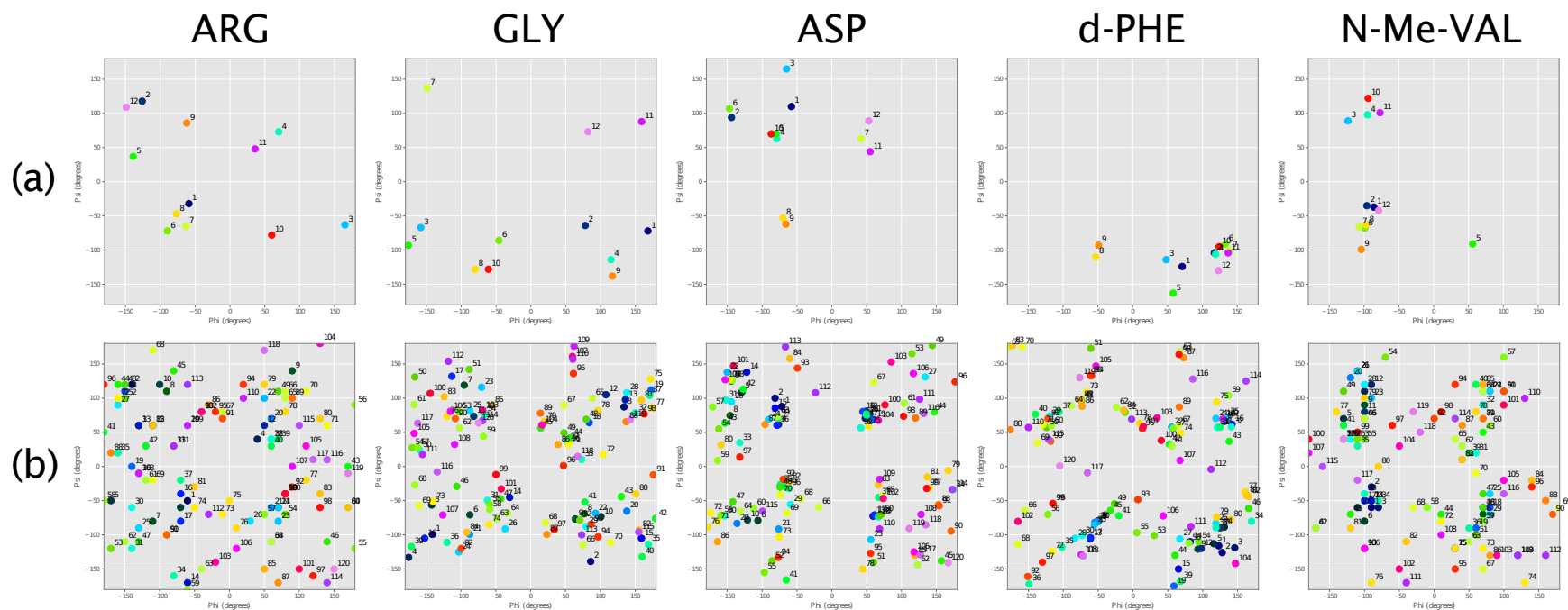


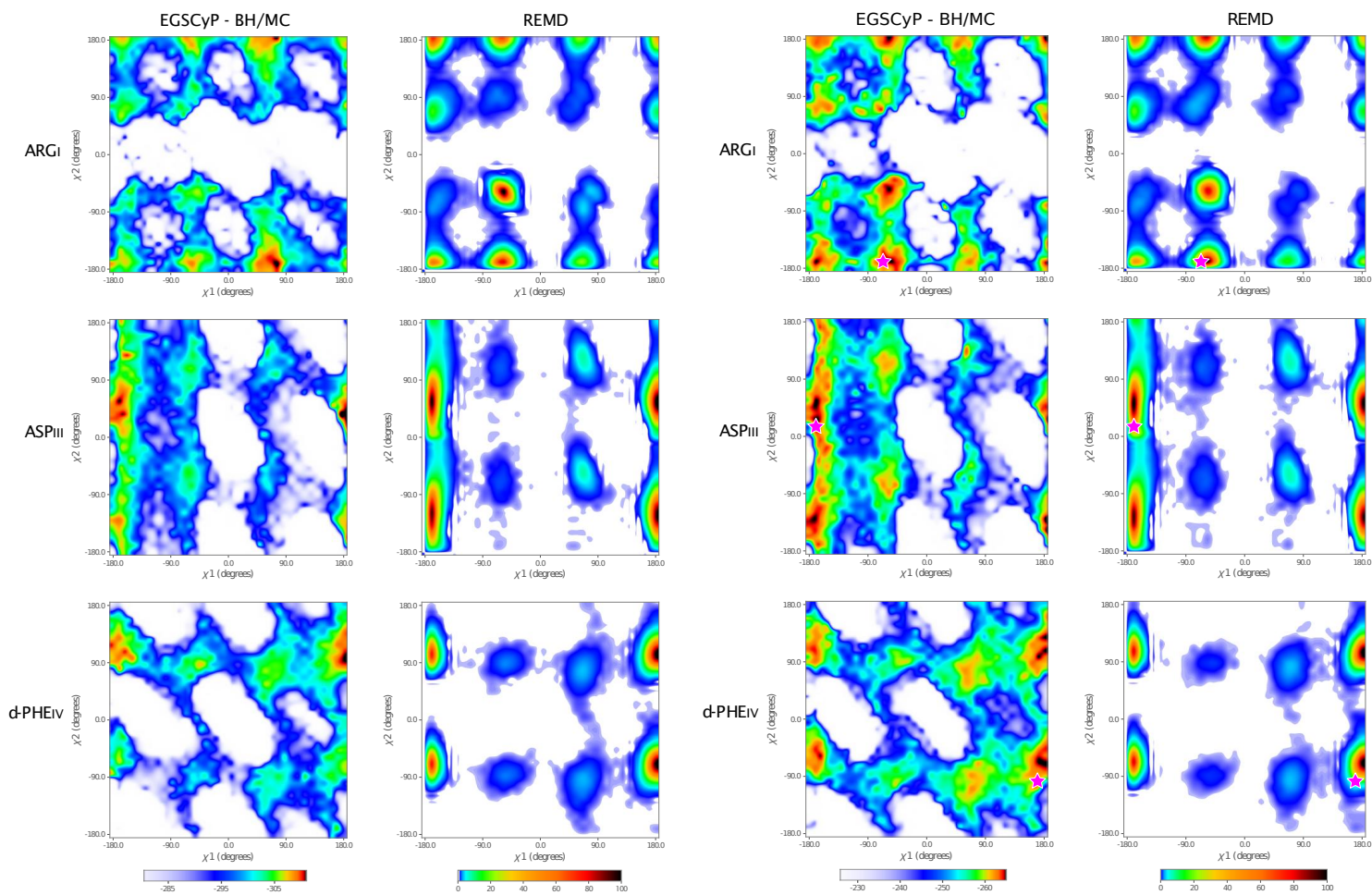
FIGURE S7 – Ramachandran diagrams, with the values of the dihedral angles for each residue (column) projected for each cluster of (a) REMD simulations, and (b) EGSCyP. Each point corresponds to a representative conformation of each cluster, with the number corresponding to the cluster number.

6 χ diagrams for RGDfV and RGDfV'

To compare the side chain sampling between EGSCyP (with the alternation of Basin Hoppin/Monte Carlo Minimizer) and REMD simulations, a projection of the energy landscape of the RGDfV and RGDfV' peptides as a function of the χ_1 and χ_2 dihedral angles for arginine, aspartate and phenylalanine residues is represented in Figure S8. The associated coverage percentages are reported in Table S7.

TABLE S7 – Coverage percentage of the χ_1 , χ_2 diagrams for arginine, aspartate and phenylalanine residues of RGDfV (left table) and RGDfV' (right table), for the EGSCyP method and the REMD simulations.

Method	ARG	ASP	d-PHE	Method	ARG	ASP	d-PHE
EGSCyP	74%	75%	74%	EGSCyP	68%	77%	80%
REMD	54%	49%	47%	REMD	55%	51%	45%



(a) RGDfv

(b) RGDfv'

FIGURE S8 – χ_1 , χ_2 diagrams for (a) RGDfv and (b) RGDfv'. The diagrams were created using the matplotlib python library¹ for arginine, aspartate and d-phenylalanine residues. The first column of each peptide corresponds to the results obtained with EGSCyP (alternation of BH/local minimization by Monte Carlo search) and the second column with REMD simulations. The color code corresponds to the potential energy in kcal/mol of the whole peptide for EGSCyP and to the normalized frequency ($nfreq_i = 100 - 100 \cdot \frac{maxfreq - freq_i}{maxfreq}$) of the conformations found during the simulations at the lowest temperature replica for REMD. The pink stars in (b) correspond to the value of the χ_1 , χ_2 dihedral angles of the crystallographic structure of Cilengitide (code PDB : 1L5G) in complex with integrin.

7 Performances comparison of both methods

TABLE S8 – Computational time in CPU hours (a) for EGSCyP and (b) for REMD simulations for exploring Cilengitide conformational landscape. For EGSCyP, three step sizes in the grid search of ϕ - ψ dihedral angle sampling are considered, as well as an exploration without Amber refinement (wA). The processors used for the two methods were different. For EGSCyP, the computations were made on one core on an Intel processor i7-7700 of 3.6 GHz. For REMD simulations, the computations were made on eight cores on the Occigen French national cluster with an Intel processor E5-2690v3 of 2.6 GHz. The processor used for EGSCyP is estimated 50% faster than the one used for REMD simulations.

(a) EGSCyP			(b) REMD		
Step Size	CPU hours	Conformations	Time	CPU hours	Frames
10°	188	541,297	2.4 μ s	960	240,000
10° - wA	94	541,259	600 ns	240	60,000
20°	10.5	33,606	240 ns	96	24,000
30°	2.75	6,535	24 ns	9.6	2,400

TABLE S9 – Coverage percentage of the Cilengitide conformational landscape (a) for EGSCyP and (b) for REMD simulations. For EGSCyP, three step sizes in the grid search of ϕ - ψ dihedral angle sampling are considered, as well as an exploration without Amber refinement (wA). For REMD, the whole trajectory was simply divided into the wanted sub-time and the percentages were computed as the mean of all the percentages of the different sub-trajectories. For instance, for the last line of (b), one hundred coverage percentages have been computed on the whole set of sub-trajectories. Their mean value is reported. For both (a) and (b), a grid spacing of 10° is used for the coverage calculation, independently of the step size or the time. This explains in (a) the large decrease in the level of coverage for the two sampled residues (ARG and VAL) with step sizes of 20° and 30°.

(a) EGSCyP					
Step Size	ARG	GLY	ASP	PHE	VAL
10°	97%	92%	87%	88%	96%
10° - wA	95%	96%	95%	93%	94%
20°	40%	82%	78%	81%	40%
30°	16%	70%	65%	69%	16%

(b) REMD					
Time	ARG	GLY	ASP	PHE	VAL
2.4 μ s	49%	45%	25%	16%	18%
600 ns	39%	36%	19%	12%	15%
240 ns	33%	32%	17%	10%	13%
24 ns	20%	22%	11%	7%	8%

TABLE S10 – Time repartition between the different steps of the EGSCyP algorithm.

Step	10°	20°	30°	Mean
IK	2%	2%	1.5%	1.83%
BH	0.9%	0.7%	1.4%	1%
MC	22%	25%	18%	21.7%
collisions	0.04%	0.04%	0.03%	0.04%
Amber	50%	50%	36%	45.33%
Other	25.06%	22.26%	43.07%	30.1%

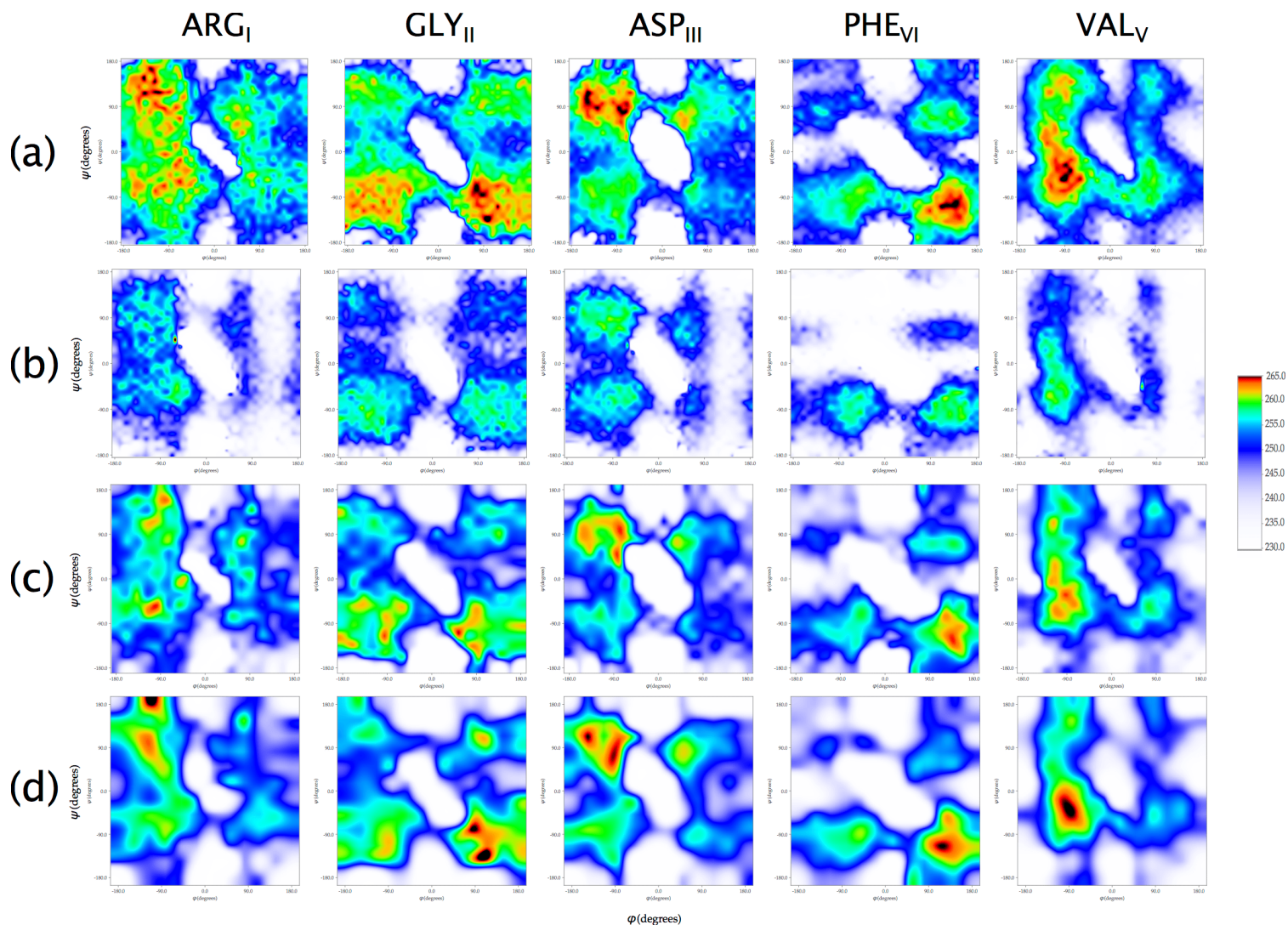


FIGURE S9 – Ramachandran plots presenting the results obtained with the EGSCyP method for Cilentide (RGDfV') with a grid search step size of (a) 10°, (b) 10° without Amber minimization, (c) 20° and (d) 30°. The diagrams were created using the matplotlib python library¹. The energy landscape was projected as a function of the dihedral angles ϕ and ψ for each residue (on each line). The color code corresponds to the minimal potential energy in kcal/mol of the whole peptide associated to each combination of ϕ - ψ angles. Energy minima are represented in black/red, while areas with no conformation are in white, and areas of high energy are in blue. The color scale was made to be unique between all diagrams. For (c) and (d), the grid spacing is 20° and 30° respectively, in order to be equal to the step size.

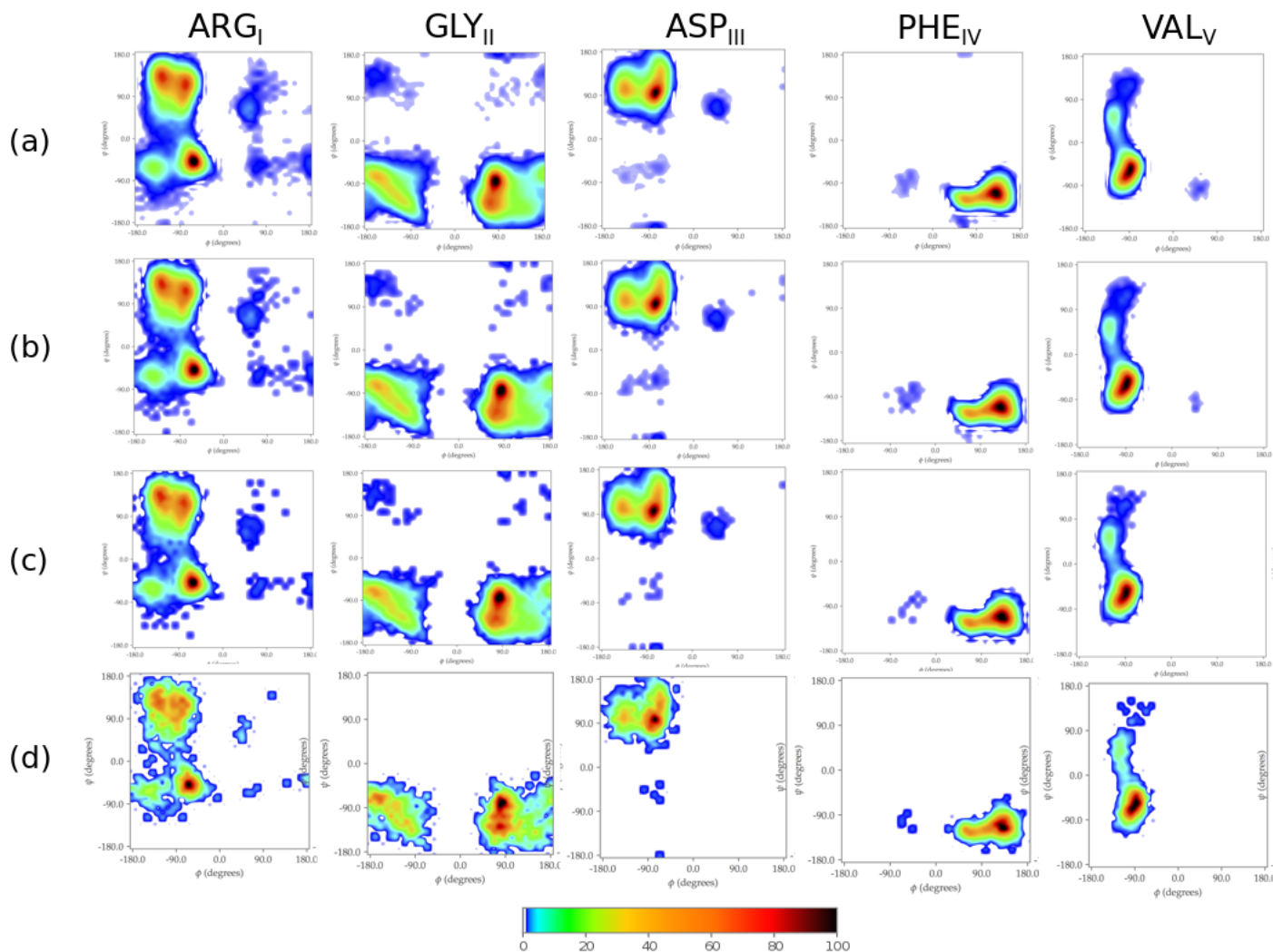


FIGURE S10 – Ramachandran plots presenting the results obtained with the REMD simulations for Cilengitide (RGDFV) with simulation time of (a) 2.4 μ s, (b) 600 ns, (c) 240 ns and (d) 24 ns. This was intended to check the run time performance of REMD, and decide among two opposite options. 1) Does it start by high energy conformations description, then refine them to low energy? 2) Does it find, even in the very first steps, some of the minima? Different simulation times were obtained by dividing the whole trajectory of 2.4 μ s into several sub-trajectories. The conformations explored during the first sub-trajectory of each division, were reported on Ramachandran maps using the matplotlib python library¹. The color code corresponds to the normalized frequency ($nfreq_i = 100 - 100 \frac{max_{freq} - freq_i}{max_{freq}}$) of the conformations found during the simulations at the lowest temperature replica is projected as a function of ϕ and ψ . The maximal frequency is thus in black and a null frequency in white. The smoothing procedure overestimates the level of coverage (for the details see Figure S11).

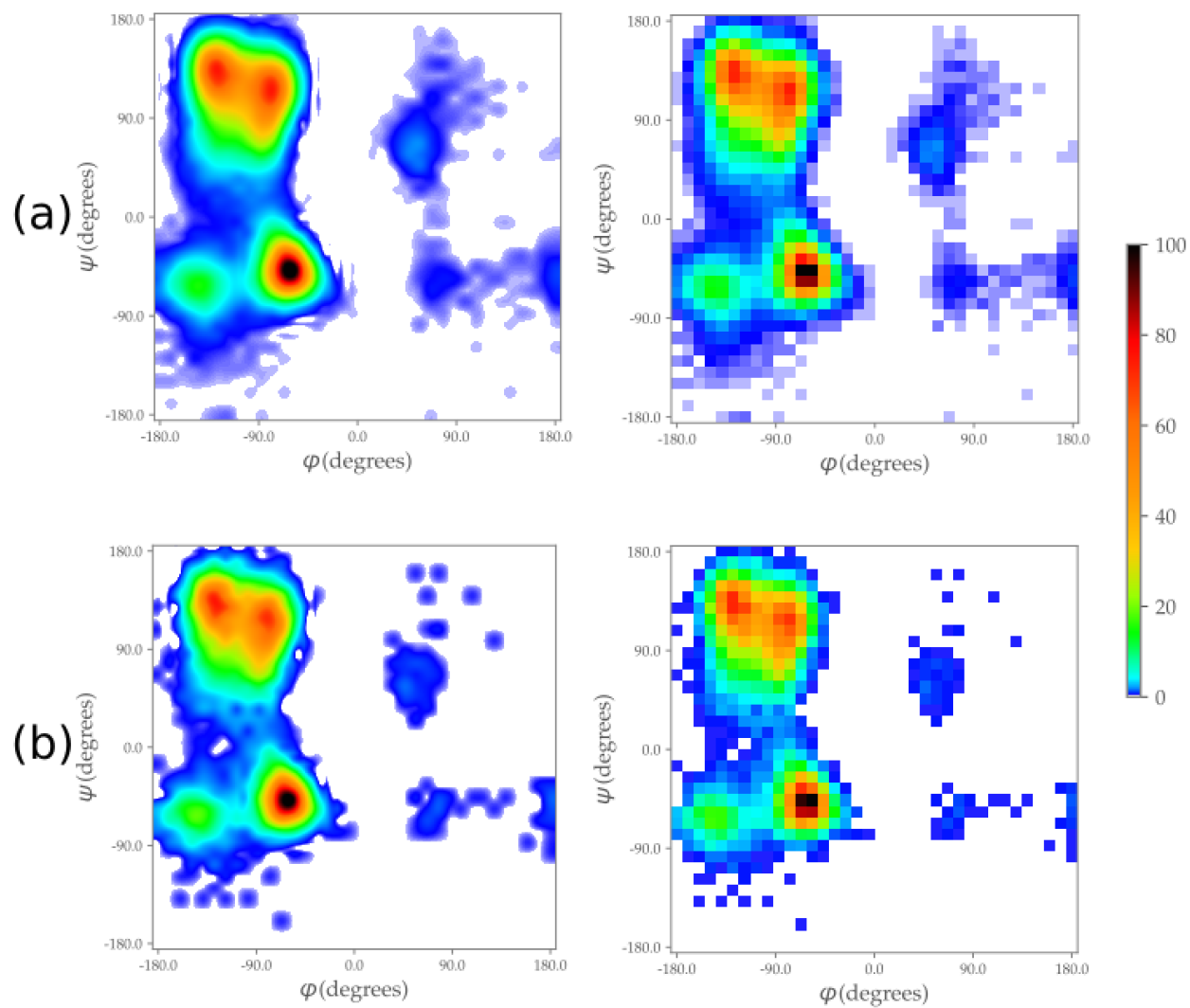


FIGURE S11 – Ramachandran plot for the ARG_I at (a) $2.4 \mu\text{s}$ and (b) 240 ns from Figure S10c with (left column) and without (right column) smoothing. For (b), the level of coverage is 33% in both maps as shown in Table S9, although it appears to be slightly overestimated on the smoothed map.

8 Superimposition of Cilengitide minima with crystallographic and NMR structures

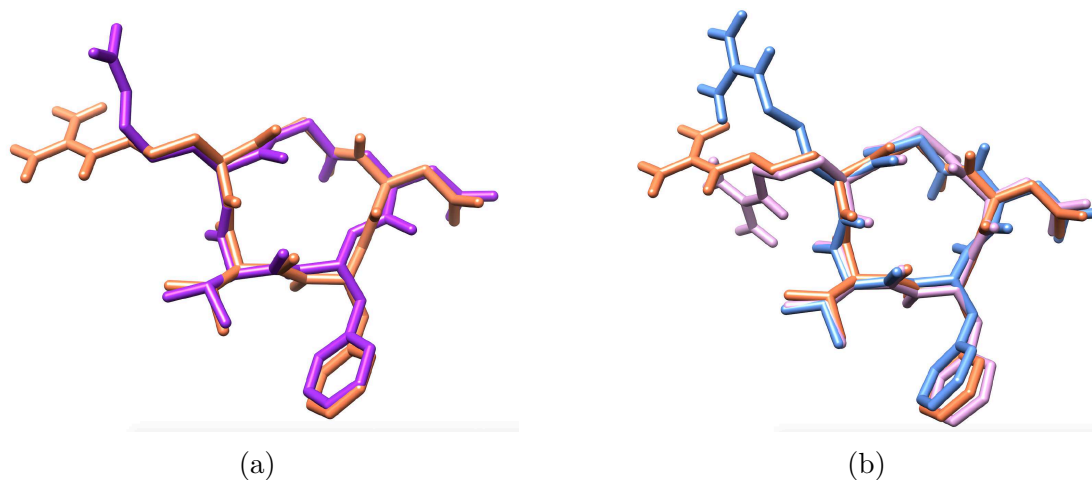


FIGURE S12 – Superimposition of experimental structures and predicted minima for Cilengitide. (a) The bound crystallographic (purple; code PDB = 1L5G) and the free NMR (coral) structures were superimposed. This last was kindly provided by Pr Horst Kessler. RMSD between these structures equals to 29° . (b) The minima $\text{min}_2^{\text{EGSCyP}}$ (blue) and $\text{min}_2^{\text{REMD}}$ (pink) were superimposed to the free NMR structure (coral) of Cilengitide. RMSD are respectively equal to 24° and 23° .

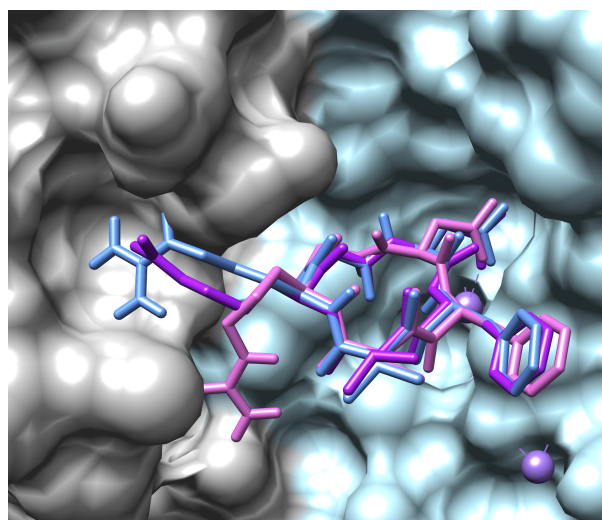


FIGURE S13 – Visualization of Cilengitide crystallographic structure (purple; code PDB 1L5G), in complex with integrin proteins (light grey and light blue in representation surface). Structures from the minima $\text{min}_2^{\text{EGSCyP}}$ (blue) and $\text{min}_2^{\text{REMD}}$ (pink) were superimposed to the crystallographic one. Lavender spheres are manganese atoms from the crystallographic structure.

Références

1. J. D. Hunter, Computing in science & engineering **9**, 90 (2007).