

## SUPPLEMENTARY INFORMATION

### Multi-scale Modeling of Mycosubtilin Lipopeptides at the air/water

#### Interface : Structure and Optical Second Harmonic Generation.

**Claire Loison**<sup>1</sup>✉, **Mehmet Nail Nasir**<sup>2</sup>, **Emmanuel Benichou**<sup>1</sup>, **Françoise Besson**<sup>2</sup> and **Pierre-François Brevet**<sup>1</sup>

<sup>1</sup> *Intitut Lumière Matière, UMR CNRS 5306, Université Claude Bernard Lyon 1, 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne cedex, France.*

<sup>2</sup> *Institut de Chimie et Biochimie Moléculaires et Supramoléculaires, UMR CNRS 5246, Université Claude Bernard Lyon 1, 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne cedex, France.*

#### **S1 All Atom Molecular Dynamics Simulations**

##### **S1a. Force-Fields and dynamics**

The simulations used the CHARMM22-CMAP force field with torsional cross-terms [1,2] for proteins, CHARMM27 [3] for the lipidic tails (taken from the phospholipids values) and water (TIP3P). The beta-amino acid was modeled using the parameters developed recently by Cui and coworkers [4]. Minimization and dynamics were conducted with the NAMD package (version 2.7b or 2.8) [5]. Particle-mesh Ewald was used to calculate the full long-range electrostatic interactions ; a cutoff of 1.2 nm was used for Lennard-Jones potentials, with a smooth switching function starting at 1.0 nm. Simulations were conducted at a constant temperature of 300 K with a damping coefficient of 5 ps<sup>-1</sup>. Periodic boundary conditions were applied in the three directions. The systems were simulated with a fixed bounding box. Bonds

involving hydrogen atoms were constrained to their equilibrium length using the SHAKE/RATTLE algorithm. Multiple-timestep integration was carried out using r-RESPA with a base timestep of 2 fs and a secondary timestep of 4 fs for long-range interactions.

### **S1b. Replica Exchange Molecular dynamics of isolated Mycosubtilin in implicit water**

Several initial structures of mycosubtilin were used for the simulation of mycosubtilin at the air/water interface. They were produced by Replica Exchange Molecular Dynamics Simulation [6,7] as implemented in NAMD2.7b. To permit a broader conformation sampling, 40 replicas between 60 K and 2400 K (exponential distribution of temperatures) have been used, with exchange attempts every 1 ps. The duration of the REMD trajectories was 50 ns for every temperature. The temperature exchange probabilities upon the whole temperature range averaged between 0.3 and 0.5. This REMD simulation was done using Generalized Born Implicit Solvent model (GBIS) as implemented in NAMD [8]. This approximation permits to calculate the electrostatic interaction between atoms in a dielectric environment described by the Poisson-Boltzmann equation. Solubilized atoms are treated as spheres of low protein dielectric ( $\epsilon_p = 1$ ), whose radius is the van der Waals radius, in a continuum of high solvent dielectric ( $\epsilon_s = 78.5$ ). For this calculation only, electrostatic interactions were neglected at long range, with cutoffs at 18 Å, with a continuous switch after 16 Å; default parameters were used for the solute surface calculations. Finally, the 50 ns-trajectory at 60 K was clusterized using the root mean square distances (RMSD) of the backbone atoms using the clustering algorithm of Daura et al. [9] was used, as implemented in the `g_cluster` tool of the GROMACS package [10] with a cutoff of 0.8 nm. This resulted in nine structures of Mycosubtilin with significantly different backbones (RMSD > 1.5 Å). The 50 ns-trajectory at 300 K was also clusterized using the root mean square distances (RMSD) of the backbone

atoms, which resulted in about 100 structures of mycosubtilin with significantly different backbones (RMSD > 1.5 Å).

#### **S1c. AA-MD of isolated Mycosubtilin in explicit water**

The 9 structures obtained as described in Section S1b have been solubilized in a cubic box of TIP3P water. The root mean square distance of the backbone atoms relative to the initial conformation were measured during 20 ns to gain insight into the impact of the solvent model on the conformation. The AA-MD distributions of bond length, angle and dihedral angles used to calibrate the CG-MD (see Section S2) were extracted from these 9 simulations of 20ns, ie 180ns of total simulations. The nine conformations obtained are added as supplementary material in the archive file : mycosubtiline\_in\_bulk\_water\_nowater.tar.

#### **S1d. AA-MD of isolated Mycosubtilin at the air/water interface**

The nine structures obtained after the 20 ns simulations as described in Section S1c have been placed at the vacuum/water interface of a slab of 3 nm of TIP3P water molecules, of area 16 nm<sup>2</sup>. The peptidic head was placed in the water, whereas the lipidic tail was placed in vacuum. The root mean square distance of the backbone atoms relative to the initial conformation were measured during 2ns.

#### **S1e. AA-MD of 25 Mycosubtilin molecules at the air/water interface**

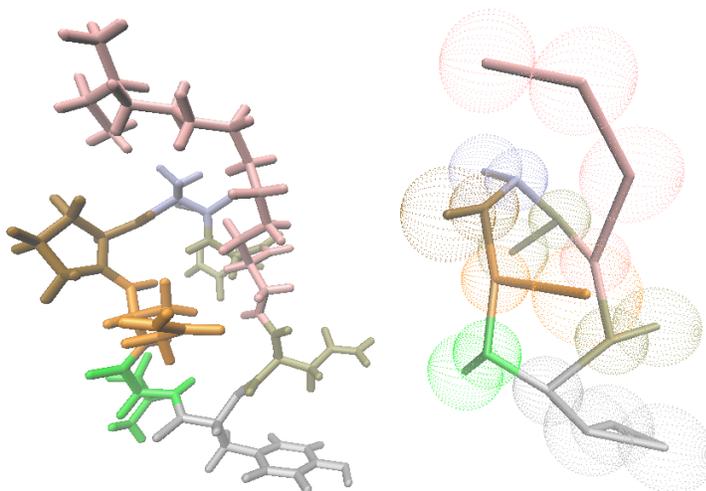
The nine structures obtained after the simulations described in Section S1d have been placed at the vacuum/water interface of a slab of 3 nm of TIP3P water molecules, of area 100 nm<sup>2</sup>. The molecules were placed on a regular cubic grid. The time evolution of the system was followed for 75 ns at 300K.

## S2 Coarse-Grained Molecular Dynamics Simulations

To avoid the restrictions on both time and spatial scales imposed by the use of the detailed force fields in molecular dynamics simulations, coarse-grained (CG) models have been used.

### S2a Force-Fields and dynamics

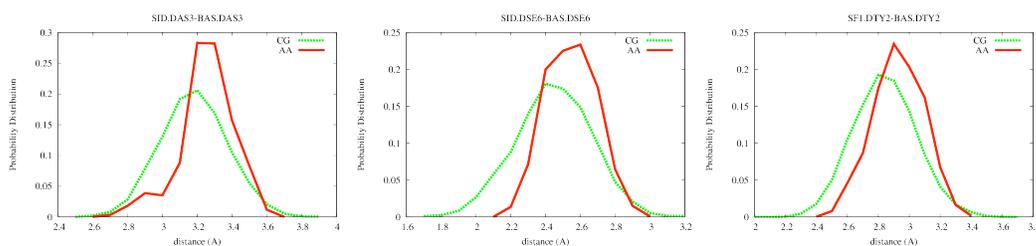
We developed a coarse-grained (CG) model of Mycosubtilin based on the MARTINI force field for biomolecules [11,12] and its extension to proteins [13]. In this model, each protein residue is represented by one backbone bead and usually one or a few sidechain beads. Each beads represent two to four heavier atoms and the hydrogens attached to it. The parameters for the side chains have been optimized to reproduce the potential of mean force to insert the amino acid sidechain in a phospholipid bilayer obtained with all-atom models [14].



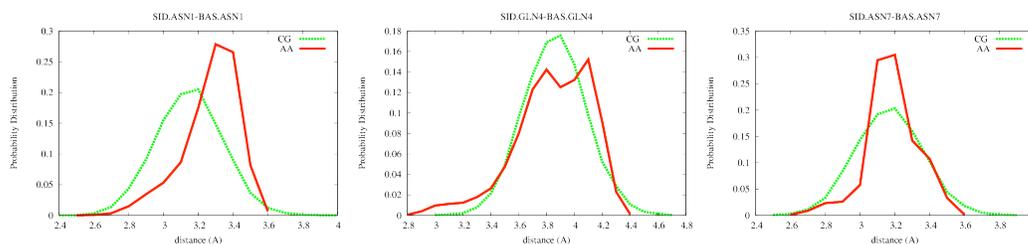
**Figure S1:** Schematic view of the all-atom model of Mycosubtilin (left) and of the MARTINI-based model (right). Each amino acid has a different color.

The original MARTINI parameters for the bonded interactions were inspired from the distributions of bond lengths and angles of the protein structure included in the Protein Data Bank (PDB) [13]. For Mycosubtilin, a few missing parameters for bonded interactions were calibrated on the all atom simulations detailed in Section S1d).

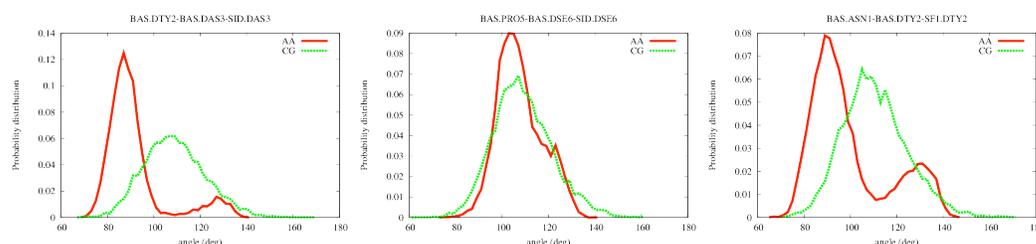
(1) The parameters for the D-amino-acids were chosen equal to the parameters for the L-amino-acids and not changed from the original MARTINI parameters. Figures S2, S3 compare the distance probability distributions in the AA and CG simulations for D and L amino acids respectively; whereas Figures S4 and S5 compare the angular distributions. The distributions of bond lengths and angles obtained with the AA model involving D-amino acids and L-amino acids were similar, as can be seen for example comparing Figure S2 and Figure S3 for the D-Asn3 and the L-Asn1 or L-Asn7. The MARTINI parameters yield BS distance distributions, which are in relatively good agreement with the AA simulations. For the BBS angles, the matching of the CG model on the AA is less satisfactory: the original MARTINI parameters yield a quasi-systematic shift of about  $30^\circ$ , but no reoptimization of the original Martini parameters were performed.



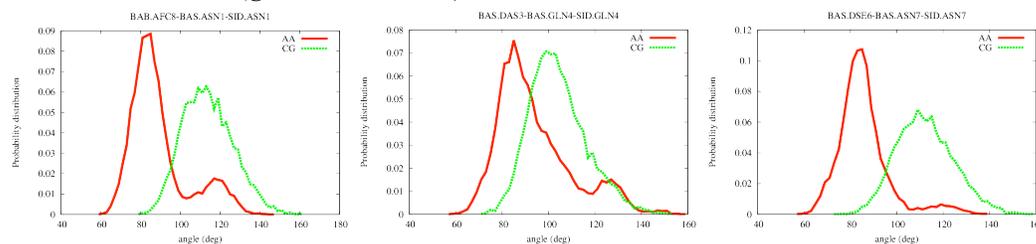
**Figure S2:** Probability distributions of the length BS obtained for the D-Asn3, D-Ser6 and D-Tyr2 amino acids in the all-atom simulations of Section S1c (red solid line) and in the CG simulations (green dashed line).



**Figure S3:** Probability distributions of the length  $BS$  obtained for the *L*-Asn1, *L*-Gln4 and *L*-Asn7 amino acids in the all-atom simulations of Section S1c (red solid line) and in the CG simulations (green dashed line).



**Figure S4:** Probability distributions of the angle  $BBS$  obtained for the *D* amino acids (*D*-Asn3, *D*-Ser6 and *D*-Tyr2) in the all atom simulations of Section S1c (red solid line) and in the CG simulations (green dashed line).



**Figure S5:** Probability distributions of the angle  $BBS$  obtained for the *L* amino acids (*L*-Asn1, *L*-Gln4 and *L*-Asn7) in the all atom simulations of Section S1c (red solid line) and in the CG simulations (green dashed line).

(2) The  $\beta$ -amino acid of Mycosubtilin was parametrized with one single bead for the backbone atoms, and one bead for four carbon atoms of the lipidic tail. The parameters of the lipidic tails equal those of the lipidic tails of phospholipids or fatty chains in the MARTINI force field [13].

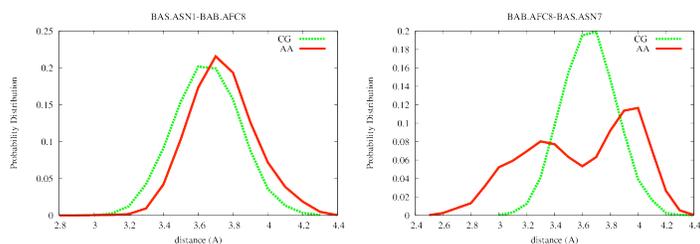
Bead	Martini Type	Charge	Mass (au)
BAB	P6C	0	72
SF1	C1	0	72
SF2	C1	0	72
SF3	C1	0	72

The non bonded parameters for the new backbone type P6C was the same as the one of P5C.

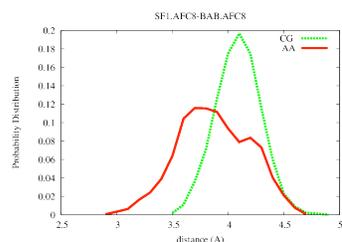
The bonded parameters were chosen as listed in Table S1.

Bond (a)	$r_0$ (nm)	k (kJ/mol/nm <sup>2</sup> )	See Figure
P6C-P5C	0.38	3722	S6
P6C-C1	0.40	1255	S7
Dihedral (b)	$\theta_0$ (°)	k (kJ/mol/rad <sup>2</sup> )	
P5C-P6C-P5C	125	20.8	S8 (left)
P5C-P5C-P6C	135	20.8	S8 (middle)
P5N-P5C-P6C	120	20.8	S8 (right)
P5C-P6C-C1	85	12.5	S9
P6C-C1-C1	180	12.5	

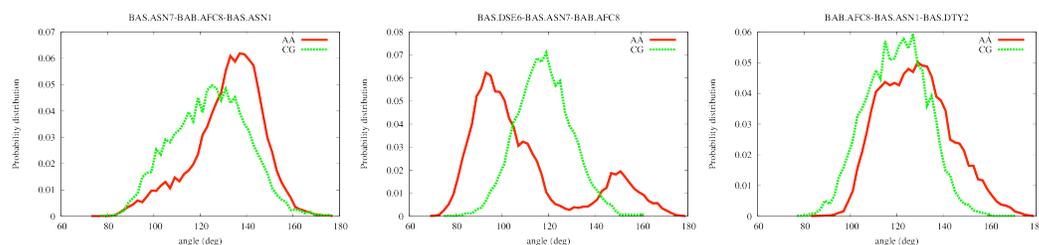
**Table S1:** New parameters for the  $\beta$ -amino-acid of Mycosubtilin. (a) In NAMD, using the MARTINI option, the bond energy is defined as  $k(r-r_0)^2$  (without factor 1/2). (b) in MARTINI, the dihedral angles are defined with cosine functions, see [13].



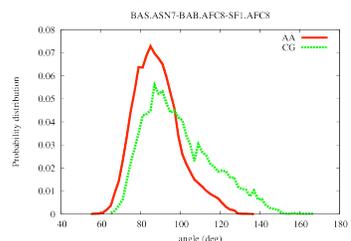
**Figure S6:** Probability distributions of the bond length  $BB$  obtained for the two neighbours of the  $\beta$ -aminoacid in the all atom simulations of Section S1c (red solid line) and in the CG simulations (green dashed line). Both covalent bonds were of the same type in the MARTINI model (P6C-P5C), whereas the AA distributions were rather different, so that a compromise was chosen.



**Figure S7:** Probability distributions of the bond length  $BS$  obtained for the  $\beta$ -aminoacid in the all atom simulations of Section S1c (red solid line) and in the CG simulations (green dashed line).



**Figure S8:** Probability distributions of the bond angle  $BBB$  obtained for the  $\beta$ -aminoacid in the all atom simulations of Section S1c (red solid line) and in the CG simulations (green dashed line).



**Figure S9:** Probability distributions of the bond angle  $BBS$  obtained for the  $\beta$ -aminoacid in the all atom simulations of Section S1c (red solid line) and in the CG simulations (green dashed line).

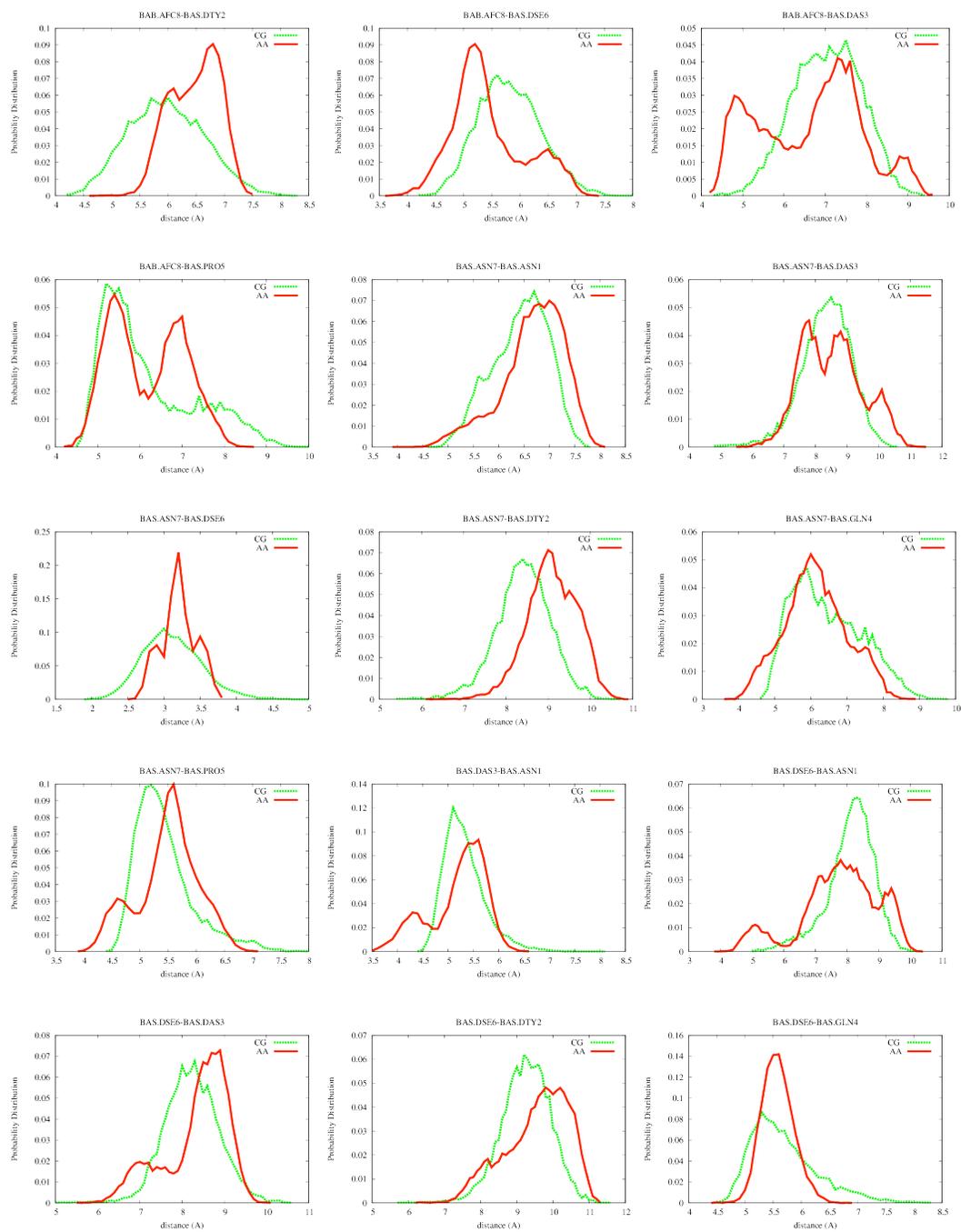
(3) All the backbone beads parameters were those of a coiled conformation (without BBBB dihedral constraints), but the structure of the backbone was maintained similar to those obtained from the all-atom simulations by additional harmonic bonds between pairs of backbone beads [15]. The strength of the constraints is very weak compared to the values typically used by Marrink and coworkers [15], but it was observed to influence nevertheless the distance probability distributions. For the sake of simplicity and in the spirit of the MARTINI force field, the strength of this constraint was chosen to be unique.  $K$  is in kcal/mol/rad<sup>2</sup> and  $L_0$  in Å.

B-B	AFA8-DAS3	AFA8-DSE6	AFA8-DTY2	AFA8-DTY2	AFA8-PRO5	ASN7-ASN1	ASN7-DAS3	ASN7-DAS3	ASN7-GLN4	ASN7-PRO5	DAS3-ASN1
$k$	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
$L_0$	7.5	5.1	6.5	6.0	6.0	6.8	8.5	9.0	6.0	5.5	5.5

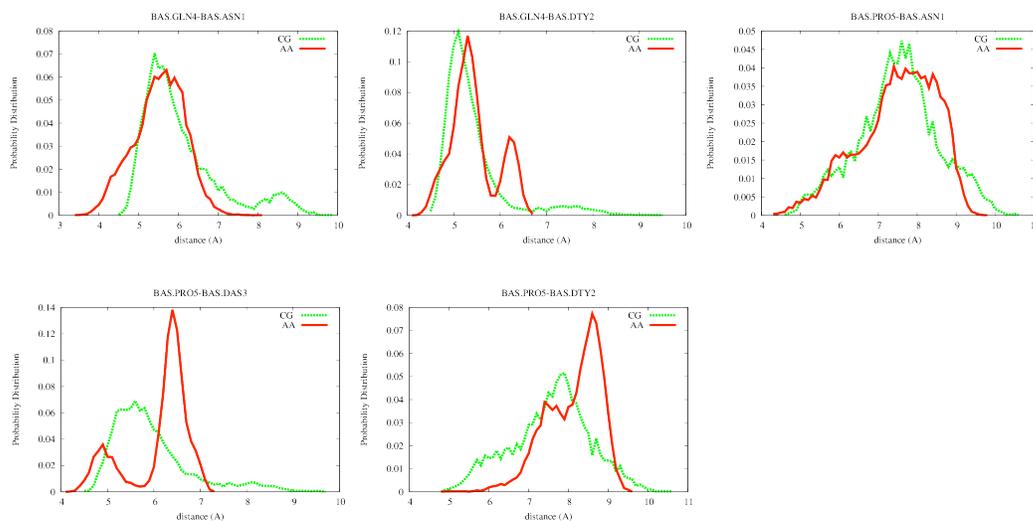
**Table S2:** List of the additional harmonic springs added to maintain the secondary structure of the peptide backbone of the CG model similar to the structure obtained with all-atom MD.

B-B	DSE6-ASN1	DSE6-DTY2	DSE6-DAS3	DSE6-GLN4	GLN4-ASN1	GLN4-DTY2	PRO5-ASN1	PRO5-DTY2	PRO5-DAS3
$k$	0.1	0.1	0.1	4.0	0.1	0.1	0.1	0.1	2.0
$L_0$	8.0	9.2	8.8	5.6	5.8	5.4	7.8	9.0	6.3

**Table S2 (continued):** List of the additional harmonic springs added to maintain the secondary structure of the peptide backbone of the CG model similar to the structure obtained with all-atom MD.



**Figure S8:** Probability distributions of some distances between backbone beads obtained for the all atom simulations of Section S1c (red solid line) and in the CG model (green dashed line).



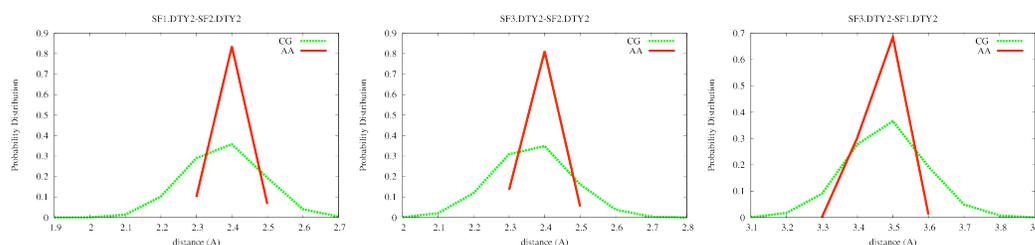
**Figure S8 (continued):** Probability distributions of some distances between backbone beads obtained for the all atom simulations of Section S1c (red solid line) and in the CG model (green dashed line).

4) In the original MARTINI force field, the tyrosine side chain is represented by an isosceles triangle, which is not coherent with the all-atom simulations. To get a precise description of the direction of the C-OH bond, the tyrosine parameters were reparametrized using a new bead type SC5y. The aim of the re-parametrization was to impose a new bond SC5y-SP1y which is constrained to remain parallel to the C-OH bond, so that we could interpret the angular distribution of the SC5y-SP1y as the angular distribution of the C-OH bond. This was allowed by the stiffness of the tyrosine aromatic moiety. The non-bonded parameters of the new bead type SC5y were identical to the previous bead type SC4y. The minor modifications of the bonded parameters (see Table S3) were supposed to have only a minor impact on the hydrophobicity and the polarity of the tyrosine side chain, which emerge from the non-bonded parameters.

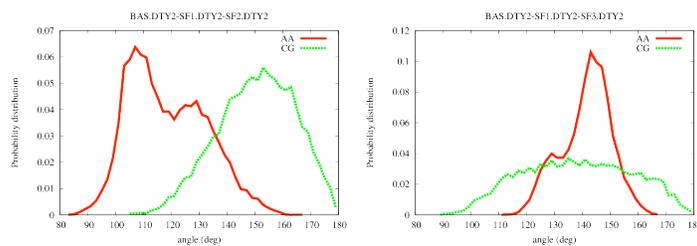
Bond	$r_0$ (nm) new	$r_0$ (nm) original	$k$ (kJ/mol/nm <sup>2</sup> ) new	$k$ (kJ/mol/nm <sup>2</sup> ) original
------	----------------	---------------------	-----------------------------------	--

P5C-SC4y	0.32	0.32	2 500	2 500
SC4y-SC5y	0.244	0.27	12 552 (*)	$\infty$
SC5y-SP1y	0.245	0.27	12 552 (*)	$\infty$
SC54-SP1y	0.355	0.27	12 552 (*)	$\infty$
<b>Dihedral (b)</b>	$\theta_0$ (°) new	$\theta_0$ (°) original	<b>k (kJ/mol/rad<sup>2</sup>)new</b>	<b>k (kJ/mol/rad<sup>2</sup>) original</b>
P5C-SC4y-SC5y	120	150	2.98	5.97
P5C-SC4y-SP1	145	150	2.98	5.97

**Table S3:** List of the old and new bond parameters for the tyrosine, including the new bead type SC5y. (\*) The high constant mimicks the constraint imposed in the original MARTINI parameters. Higher values impede the use of the large time step (20 fs).



**Figure S9:** Probability distributions of the bond lengths obtained for the tyrosin side chain from the all atom simulations of Section S1c (red solid line) and in the CG simulations (green dashed line).



**Figure S10:** Probability distributions of the angles obtained for the tyrosin side chain in the all atom simulations of Section S1c (red solid line) and in the CG simulations (green dashed line).

## S2b. Set up and dynamics

One of the structure obtained after the simulations described in Section S1d has been placed at the vacuum/water interface of a slab of 3 nm of W (and WAF) MARTINI water molecules. Two systems were built : 12 Mycosubtilins on a water surface of 57 nm<sup>2</sup> (4.7 nm<sup>2</sup> per lipid) and 121 Mycosubtilins on a water surface of 121 nm<sup>2</sup> (1 nm<sup>2</sup> per lipid). The simulation parameters used the MARTINI standard settings for the nonbonded interactions [12] : the dielectric constant is 15, a cutoff of 1.2 nm ( $r_{\text{cut}}$ ) was used in the calculation of nonbonded interactions with a shifted function. The Lennard-Jones potential is shifted from 0.9 to 1.2 nm. The electrostatic potential is shifted from 0.0 to 1.2 nm. Both the energy and the force vanish at the cutoff distance. Systems were simulated in the canonical ensemble, at constant particle number, volume, and temperature (NVT ensemble) using periodic boundary conditions in the three dimensions. The temperature was kept constant by coupling to a Langevin heat bath at 300 K ( $\tau_T = 0.2$  ps). The timestep was 20 fs. The system with 12 Mycosubtilins was equilibrated during 1  $\mu$ s and data was accumulated for 2  $\mu$ s. The system with 121 Mycosubtilin was equilibrated during 4  $\mu$ s, and data was then accumulated for 2  $\mu$ s. Input setup and analysis was performed using VMD and topotools package. Minimization and dynamics of the CG model were conducted with the NAMD2.8 package.

## REFERENCES

- [1] A.D. MacKerell *et al.*, *J. Phys. Chem. B*, **102** (1998) 3586
- [2] A.D. MacKerell, M. Feig, C.L. Brooks, *J. Am. Chem. Soc.*, **126** (2004) 698
- [3] S.E. Feller, A.D. MacKerell, *J. Phys. Chem. B*, **104** (2000) 7510
- [4] X. Zhu, P. Koenig, M. Hoffmann, A. Yethiraj, Q. Cui, *J. Comput. Chem*, **31** (2010) 2063
- [5] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot,

- R.D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.*, **26** (2005) 1781
- [6] A. Patriksson, D. Spoel, *Phys. Chem. Chem. Phys.* **10** (2008) 2073
- [7] Y. Sugita, A. Kitao, Y. Okamoto, *J. Chem. Phys.* **113** (2000) 6042
- [8] D. E. Tanner, K.-Y. Chan, J. C. Phillips, K. Schulten, *J. Chem. Theory Comput.*, **7** (2011) 3635
- [9] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark  
*Angew. Chem. Int. Ed.* **38** (1999) 236
- [10] B. Hess, C. Kutzner, D. Van der Spoel, E. Lindhal, *J. Chem. Theory Comput.*, **4** (2008) 435
- [11] S.J. Marrink, A.H. de Vries, A.E. Mark, *J. Phys. Chem. B*, **108** (2004) 2750-
- [12] S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, and A.H. de Vries *J. Phys. Chem. B* **111**(2007), 7812
- [13] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, S.-J. Marrink  
*J. Chem. Theory Comput.*, **4** (2008) 819
- [14] J.L. MacCallum, W.F.D. Bennett, D.P. Tieleman, *J. Genet. Physiol.*, **129** (2007) 371
- [15] X. Periole, M. Cavalli, S.J. Marrink, M.A. Ceruso, *J. Chem. Theory Comput.*, **5** (2009) 2531