Supporting Information

Role of hydration in collagen recognition by bacterial adhesin

Luigi Vitagliano¹,*, Rita Berisio¹, Alfonso De Simone²,*

¹Istituto di Biostrutture e Bioimmagini, CNR, Via Mezzocannone 16, I-80134 Naples, Italy
²Department of Chemistry, University of Cambridge, Lensfield Road CB2 1EW, Cambridge UK.
*Email: ad491@cam.ac.uk or luigi.vitagliano@unina.it ;
Supplementary Methods

Simulations Parameters
The Berendsen algorithm was applied for the temperature and pressure coupling. The bonds were constrained by the LINCS algorithm. The particle-mesh Ewald (PME) method was used to account for the electrostatic contribution to non-bonded interactions (grid spacing of 0.12 nm). To ensure a system at pH 7 the protonation states of pH-sensitive residues were as follows: Arg and Lys were positively charged, Asp and Glu were negatively charged, and His was neutral. In line with a recent study (1), whose side chain stabilizes through hydrogen bonding interactions the isopeptide between Lys 176 and Asp 293, was protonated. The net charge of the protein was neutralized by the addition of Cl- and Na+ ions. The set of parameters recently developed by Pande and coworkers have been used for Hyp residues (2). The calculations were extended for a simulation time of 50ns in each system. The evolution of structural parameters were checked by using GROMACS analyses routines and in-house programs. We checked that in all simulations, the minimum distance between the system and its images generated by the periodic boundary conditions was always larger than the cut-off values used for non-bonded interactions.

Solvent Density Maps
For each frame of the sampling, water molecules positions were counted in a grid of 0.5 Å after superimposing the current protein structure onto a reference one. To prevent sweeping effects due to backbone flexibility, trajectory frames have been selected based on their Cα-RMSD with a reference structure. In particular, only structures with a Cα-RMSD value lower than 1.0 Å from the reference conformation were considered. To maximize the efficiency of the sampling, the reference structure has been selected with a clustering method based on Cα-RMSD. The maps were stored in EDM format and drawn by means of curve levels. When the molecular fluctuations are dominated, as in this case, by the mutual intermotion of otherwise rather rigid domains, the global alignment allows producing hydration maps with similar accuracy as maps produced by using a local superimposition. This is exemplified in figure S1 where the maps generated on the isolated domains of CNA result in high agreement with that generated by considering the whole protein despite in the first case ~98% of the conformations of the sampling were selected for having local Cα-RMSD values lower than 1.0 Å whereas in the second case only ~10% of the conformations showed global Cα-RMSD values within the limit of 1.0 Å. The advantage of using a global alignment protocol is that it allows to generate a global hydration map from which the hydration properties of the two domains can be compared.

Water Residence Times in the Hydration Sites
The time autocorrelation function P(τ) was adopted to retrieve the residence time of water molecules in the MDHS. This function is accounted by the following integral:

\[ P(\tau) = \int_0^\infty \delta(W(t) - W(t + \tau)) \, dt \]  

(1)

where the delta function \( \delta(W(t), W(t + \tau)) \) assigns the value 0 if the indexes of the
partners at times $t$ and $t + \tau$ differ or 1 in the case they differ. The resulting function is then fitted with an exponential decay equation.

**Thermodynamics integration of water binding**

We here calculated the free energy contributions of removing/inserting a water molecule from/into either a protein cavity or the bulk solvent by using the thermodynamics integration technique. In particular these have been performed by using the “slow growth” method in which the Hamiltonian ($H$) of the system is perturbed from a state A to a state B in such a way that the water molecule parameters are turned off during the calculation. Thus Coulomb and Lennard-Jones interactions of the given water are linearly reduced to zero. This is achieved by formulating $H$ as a function of a coupling parameter $\lambda$, $H(p, q; \lambda)$, which can range from 0 to 1. The calculation is performed by perturbing $\lambda$ from 0, describing the starting configuration A where the water parameters are unperturbed (or zero in the reverse simulations), to a value of 1 (configuration B) where the parameters of the water are zero and the molecule is no longer interacting with any particle of the simulation (or are unperturbed in the reverse simulations). Thus the starting and final forms of the Hamiltonian are:

$$H(p, q; 0) = H_A(p, q); \quad H(p, q; 1) = H_B(p, q)$$

(2)

Where $p$ and $q$ represent the Cartesian coordinates, conjugate momenta, respectively. The derivate of the free energy respect to $\lambda$ was evaluated according to:

$$\frac{\partial G}{\partial \lambda} = \frac{\int \left( \frac{\partial H(p, q; \lambda)}{\partial \lambda} \right) d \mathcal{P} d \mathcal{Q}}{\int \exp \left( -\beta H(p, q; \lambda) \right) d \mathcal{P} d \mathcal{Q}}$$

(3)

where $\beta$ is the inverse temperature ($K_B T^{-1}$). The free energy difference was computed by integrating Eq. 3:

$$\Delta G_{AB} = G_B(p, T) - G_A(p, T) = \int \left| \frac{\partial H(p, q; \lambda)}{\partial \lambda} \right|_{NpT, \lambda} d \lambda$$

(4)

It is worth noting that, with this approach, the endpoints of these integrations usually lead to extremely noisy energy profiles when the parameters are approaching the zero value, an effect that is mainly due to fluctuations arising from the vdW component. To circumvent this problem, the endpoints of the integration were interpolated with a spline function.

**References**


Figure S1. Solvent density map of adhesin CNA calculated with global alignment and local alignment. Adhesin CNA is drawn by yellow ribbons. MDHS arising from the solvent density map are drawn by curve levels embedding regions with a density twice higher than the bulk value. Top and bottom views show the hydration of the N1 and N2 domains, respectively. Black, red and blue curve levels are used for hydration maps calculated by superimposing the simulation frames on the whole molecule, N1 domain and N2 domain, respectively.
Figure S2.
Residue Hydration Indices. A) Correlation between Dewetting Index and Average Residue Hydration. The latter is defined in unit of bulk water density. The blue lines mark the threshold for dewetted surficial residues. According to the decay curve fitting, a dewetting index higher than 0.69 corresponds to an Average Residue Hydration value lower than one, the value of the bulk. This threshold is used to define a dewetted residue. B) Correlation between number of grid points of an aaHS and its dewetting index.
Figure S3.
Evolution of structural parameters in the ApoCNA simulation: (A) secondary structure, and (B) gyration radius of the entire molecule (black) and of the individual domains (green and red). (C) RMSD values of trajectory structures of ApoCNA simulation versus the starting X-ray model computed on the Cα atoms. The RMSD values for the entire protein, the N-terminal and C-terminal domains are reported in black, red, and green, respectively.
Figure S4.
Accuracy of the MDHS of ApoCNA. The agreement between MDHS and the X-ray water sites has been checked using the crystalline structure of the protein (PDB code 2F68). The MDHS are located on the X-ray structure by means of a local superimposition. Panels A and B show waters with higher and lower accuracy than 1.4 Å, respectively. Adhesin backbone is drawn by means of green ribbons. X-ray waters are represented by red CPKs whereas the center of MDHS by yellow CPKs.
Figure S5.
Distributions of water-water hydrogen bonds residence times. The top distribution refers to residence times of hydrogen bonds between waters in the bulk (average value 9.4 ps, standard deviation 1.2 ps). The bottom distribution refers to residence times of hydrogen bonds that water from the second hydration shell make with waters from the first hydration shell (average value 15.3 ps, standard deviation 10.3 ps).
Figure S6.
Comparison between dewetting prone patches (a) and crystal contact patches (b). The firsts (a) have been selected according to the dewetting index (see methods and Figure S2). The seconds have been individuated by using crystal symmetry operations and selecting all the residues that presented a crystal contact with other proteins in the crystal. The criterion for the contact between two residues belonging to two different proteins was to show at least one interatomic distance lower than 4Å.
Figure S7. Structural parameters of the CNA_coll simulation. (a) RMSD from the X-ray model along the trajectory of the complex CNA-coll in the simulation. Color codes are black, red, green and blue for the domain N1, domain N2, the collagen triple helix and the whole complex, respectively. (b) Triple helix bending angle of the trajectory structures. This angle is calculated from the centre of masses of the N-terminal, central and C-terminal fragments of the collagen peptide. (c) Evolution of total main chain–main chain hydrogen bonds from the collagen-like triple-helix motif in the simulation of the complex (CNA_coll). (d) Time evolution of the secondary structure of the CNA protein in the simulation of the complex (CNA_coll). (e) Distribution of the global bending angles of collagen in the free (blue histogram) and bound (orange histogram) states. (f) Autocorrelation function of the global bending angles of collagen in the free (blue) and bound (black) states. The characteristic times, estimated as $\tau_2$ of a second order exponential decay function, are 65 ps and 201 ps for the free and bound states, respectively.
Figure S8.
Close-up view of key hydration sites in CNA-coll. Left panels show hydration sites in the adhesin-collagen complex. Cyan, Yellow and violet ribbons are used for domains N1, N2 and connecting loop, respectively. Collagen is drawn by red traces connecting the Cα atoms. Sidechains involved in water mediated bridges are drawn by sticks. Water hydration sites are marked by curve levels embending regions with high solvent density. A sphere is drawn at the center of the MDHS. Right panels report the comparison of the hydration sites calculated in the adhesin-collagen complex (green curve levels) and the collagen alone (red curve levels). For clearness collagen only is drawn in a similar orientation to that adopted in the respective left panels.
Figure S9:
Comparison of CNA-coll hydration patterns at different stages of the simulation. For each simulation point, the diagrams report the distribution of the distances of the water sites compared to the final ones (50 ns).