Self-association of a highly charged arginine-rich cell-penetrating peptide

Giulio Tesei[a,1], Mario Vazdar[b], Malene Ringkjøbing Jensen[c], Carolina Cragnell[a], Phil E. Mason[d], Jan Heyda[e], Marie Skepö[a], Pavel Jungwirth[d,1], and Mikael Lund[a,1]

[a]Department of Chemistry, Division of Theoretical Chemistry, Lund University, SE-22100 Lund, Sweden; [b]Division of Organic Chemistry and Biochemistry, Rudjer Bošković Institute, HR-10002 Zagreb, Croatia; [c]Univ. Grenoble Alpes, CNRS, CEA, IBS, 38000 Grenoble, France; [d]Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, 16610 Prague 6, Czech Republic; and [e]Department of Physical Chemistry, Faculty of Chemical Engineering, University of Chemistry and Technology, Prague, 16628 Prague 6, Czech Republic

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved September 7, 2017 (received for review July 11, 2017)

Small-angle X-ray scattering (SAXS) measurements reveal a striking difference in intermolecular interactions between two short highly charged peptides—deca-arginine (R10) and deca-lysine (K10). Comparison of SAXS curves at high and low salt concentration shows that R10 self-associates, while interactions between K10 chains are purely repulsive. The self-association of R10 is stronger at lower ionic strengths, indicating that the attraction between R10 molecules has an important electrostatic component. SAXS data are complemented by NMR measurements and potentials of mean force between the peptides, calculated by means of umbrella-sampling molecular dynamics (MD) simulations. All-atom MD simulations elucidate the origin of the R10–R10 attraction by providing structural information on the dimeric state. The last two C-terminal residues of R10 constitute an adhesive patch formed by stacking of the side chains of two arginine residues and by salt bridges formed between the like-charge ion pair and the C-terminal carboxyl groups. A statistical analysis of the Protein Data Bank reveals that this mode of interaction is a common feature in proteins.

cell-penetrating peptide | self-association | MD simulations | SAXS | NMR

Recent studies focusing on interactions of charged proteins in electrolyte solutions have highlighted the interplay of two counteracting electrostatic forces (1–4). The first one originates from the presence of a localized distribution of charges defining an electrostatic patch on the protein surface. Depending on relative orientations, the charge distributions in the patches on the protein molecules can become complementary, thereby leading to an attractive electrostatic force (5). This anisotropic force is short ranged and is hereafter referred to as the electrostatic adhesive force. The other force is the double-layer force arising from the Coulombic repulsion between like-charged molecules in the electrolyte medium. Both electrostatic adhesive and double-layer forces are weakened by the presence of salt in the solution. As a nontrivial consequence, the propensity of the proteins to aggregate is heightened at low-to-intermediate ionic strengths (1–3, 6). This is because of lowering of Coulombic repulsion due to salt screening of the net charge of the protein, in conjunction with the presence of the adhesive force, which operates at shorter distances and is therefore less efficiently screened.

In this work, the competition between electrostatic adhesive and double-layer forces, together with a chemically specific like-charge attraction between guanidinium (Gdm^+) side-chain groups, is reported for solutions of a small highly charged peptide. Observation of the complex aggregation behavior for a relatively simple molecule is substantiated by a comparative investigation of deca-arginine (R10) and deca-lysine (K10), in high and low ionic strength solutions, conducted using small-angle X-ray scattering (SAXS) measurements, all-atom molecular dynamics (MD) simulations, and ^1H-^{13}C heteronuclear single quantum coherence (HSQC) NMR measurements.

The choice of the peptides is motivated by the biological interest in the high cellular uptake of arginine-rich peptides (RRPs), which has been the subject of several comparative studies (7–11). Oligo-arginine chains of 6–15 aa readily translocate across cell membranes (7, 12), while the translocation efficiency of oligo-lysines of equal length is considerably lower (7, 13). If a certain threshold peptide concentration is exceeded, cell penetration of RRPs can occur in a nonendocytotic mode (14). Although the molecular details of this transduction mechanism still need to be clarified, an important feature has been identified as the Gdm^+ moiety of the arginine side chain, which can form bidentate hydrogen bonds with phosphate and glycerol groups of the lipid molecules in the cell membrane. These interactions may promote the adsorption of RRPs on the membrane surface and perturb the packing of lipids in the bilayer (10), as well as significantly increase the lifetime of transient membrane pores (15). Further, it has been proposed that transduction is enhanced by a cooperative action between RRPs (15). This occurs when the timescale of the peptide-induced kinetic stabilization of the pores is faster than the timescale of the lateral diffusion of adsorbed peptides to a transient pore. Since the coverage of the membrane surface by the peptides influences the proposed phenomenon, the aggregation of oligo-arginines may contribute to their superior uptake rate, compared with oligo-lysines. In fact, MD simulations support the tendency for oligo-arginine molecules to dimerize,

Significance

Arginine-rich cell-penetrating peptides are promising candidates for intracellular drug delivery. These cationic peptides spontaneously traverse biological membranes via a direct mode of entry which is not yet fully understood. In this study, we report the complex solution behavior of the cell-penetrating peptide deca-arginine. Despite its large net positive charge, deca-arginine self-associates at low-to-intermediate ionic strengths, owing to an interaction mode which is present in the structure of a significant number of proteins. Self-association may thus enhance the bioavailability of deca-arginine. Our findings provide the key to the understanding of the self-association mechanism in deca-arginine with implications for the potential biological roles of this unusual binding motif.

Author contributions: G.T., M.V., C.C., P.E.M., J.H., M.S., P.J., and M.L. designed research; G.T. and M.R.J. performed research; G.T. and M.R.J. analyzed data; G.T., P.J., and M.L. wrote the paper; G.T. performed SAXS experiments and MD simulations; and M.R.J. performed NMR experiments.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Datasets and Jupyter Notebooks for reproducing the analyses of SAXS data, MD simulations, and PDB structures, as well as all presented plots can be accessed at Zenodo, https://doi.org/10.5281/zenodo.825460.

[1]To whom correspondence may be addressed. Email: Mikael.Lund@teokem.lu.se, giulio.tesei@teokem.lu.se, or pavel.jungwirth@uochb.cas.cz.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1712078114/-/DCSupplemental.

as reported for R10 adsorbed on negatively charged lipid bilayers (9, 16).

In the following sections, we first present experimental findings based on SAXS measurements on solutions of R10 and K10, as well as of R8KR and K8RK molecules. The latter two molecules are mutants of R10 and K10, respectively, resulting from a ninth residue interchange. Based on the SAXS results, we infer that an attraction, which is mainly of electrostatic origin, is present between R10 molecules, while the other molecules solely repel each other. Moreover, SAXS data for R10 at low ionic strength are consistent with the presence of dimers or oligomeric forms in solution.

Subsequently, we report HSQC NMR data and potentials of mean force (PMFs) obtained from all-atom MD simulations, which support the interpretation of the scattering data and provide an atomistic description of R10 dimers. Finally, we analyze the occurrence of the inferred mode of interaction in biological systems by inspection of protein crystal structures in the Protein Data Bank (PDB).

Results

Scattering Intensity Curves. Fig. 1 shows SAXS measurements on R10 and K10 concentration series at increasing ionic strength, c_s, at 293 K and pH 7.8. The scattering intensities $I(q)$, normalized by the peptide concentration, c_p, are reported as a function of the scattering vector, q.

At $c_s = 0.060$ M and $c_s = 0.150$ M, the crowding effect on the scattering profiles of the two peptides is strikingly different. At low q values, $I(q)/c_p$ decreases with increasing c_p for K10, while it increases for R10. Scattering intensity curves obtained for K10 are as expected for polyelectrolyte solutions of low ionic strength (17, 18), displaying an increasingly pronounced maximum that shifts to higher q with increasing c_p. Maxima are also observed in $I(q)/c_p$ for R10; however, with increasing c_p the shift of the maxima to higher q is less pronounced than for K10. At $c_s = 0.300$ M, $I(q)/c_p$ at low q decreases with increasing c_p for K10, while an oscillating trend is observed for R10. Fig. 1*D* shows $I(q)/c_p$ values extrapolated to $q = 0$ (*SI Materials and Methods*) for solutions of R10, K10, R8KR, and K8RK for increasing c_p at $c_s = 0.150$ M. While for K10 and K8RK $I(0)/c_p$ values are similar to each other (as expected), the discrepancy in $I(0)/c_p$ values between R10 and R8KR is large compared with the small difference in molecular weights (*SI Materials and Methods* and Fig. S1). Finally, $I(0)/c_p$ vs. c_p has the same decreasing trend for K10, R8KR, and K8RK, while it has an increasing trend for R10 (*SI Results* and Fig. S2).

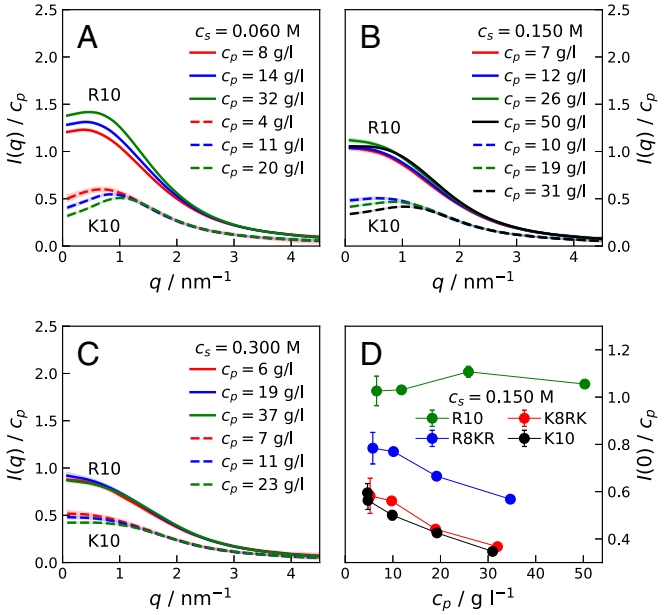

Fig. 1. (*A*–*C*) Concentration-normalized scattering intensities for R10 (solid lines) and K10 (dashed lines) at various peptide concentrations, c_p, in 0.020 M Tris buffer solutions of 0.060 M (*A*), 0.150 M (*B*), and 0.300 M (*C*) ionic strength, c_s. (*D*) Extrapolated $I(0)/c_p$ values for samples of R10, K10, R8KR, and K8RK of increasing c_p at $c_s = 0.150$ M. Lines connecting points in *D* are guides to the eye.

PMFs. Fig. 2 shows the PMFs obtained from umbrella-sampling MD simulations for pairs of R10 and K10, as a function of the intermolecular separation (*SI Materials and Methods* and Fig. S3). The intermolecular separation for R10 corresponds to the distance between the guanidino-C atoms of the ninth residues. This reaction coordinate was chosen based on inspection of the R10 dimer observed in previous MD simulations (9). Analogously, the intermolecular separation for K10 is calculated as the distance between the ϵ-C atoms of the ninth residues. At $c_s = 0.01$ M, the free energy of interaction between pairs of K10 molecules is repulsive and decays with increasing interparticle separation, as expected for two like-charged molecules in solution. In contrast, the PMF for R10 chains shows at low ionic strength a minimum at 0.4 nm, followed by two maxima at separations of 0.58 nm and 0.93 nm. As evidenced by the red and blue lines in Fig. 2, the positions of the minima are preserved at higher c_s. At $c_s = 0.07$ M, PMFs of both R10 and K10 show lower free energy values than at $c_s = 0.01$ M, corresponding to a decreased repulsion between the like-charged molecules due to electrostatic screening. At $c_s = 0.32$ M, the decay of the repulsive interaction with increasing intermolecular separation is considerably steeper than for the PMFs at lower c_s; nonetheless the free energy of the minimum at 0.4 nm separation between R10 molecules is higher compared with $c_s = 0.07$ M. At large separations, the PMFs agree with the Debye–Hückel approximation, shown by the black points in Fig. 2. The vertical error bars reflect the fluctuation of separations between the center of mass of the peptides in the umbrella-sampling simulation windows. While K10 molecules repel each other at all length scales and salt concentrations, R10 displays salt-dependent attraction at short separation. The free energy of the minimum at 0.4 nm varies nonmonotonically with c_s, and it is the lowest at $c_s = 0.07$ M. On the contrary, the difference in free energy between the maximum at 0.58 nm and the minimum at 0.4 nm decreases with increasing c_s. To facilitate further discussion of the nonmonotonic trend for the free energy minimum, Fig. 2 displays 2D schematic representations of the R10 dimeric structures.

Hydrogen Bonding in Dimeric Structures. Fig. 3 displays the probability to form an H bond between the C-terminal carboxyl group (COO^-) of the first peptide and the last five C-terminal residues of the second peptide. The H-bond probabilities are calculated from umbrella-sampling MD simulation windows, where the model peptides are at the closest separation along the reaction coordinate. It is evident for both R10 and K10 that the C-terminal COO^- of the first peptide is most likely to form H bonds with the ninth residue of the second peptide. A nonmonotonic dependence on c_s is observed for the H-bond probability between the C-terminal COO^- and the ninth residue. Specifically for R10 it increases approximately twofold from $c_s = 0.01$ M to $c_s = 0.07$ M, while at $c_s = 0.32$ M it drops back to an intermediate value. The H-bond probability between the COO^- group and the ninth residue is significantly larger for R10 than for K10. Representative snapshots from the analyzed trajectories are included in Fig. 3 to show the H-bonding pattern in the C-terminal residues of R10 and K10. In the R10 dimer, H bonds are formed between the C-terminal COO^- and the Gdm^+

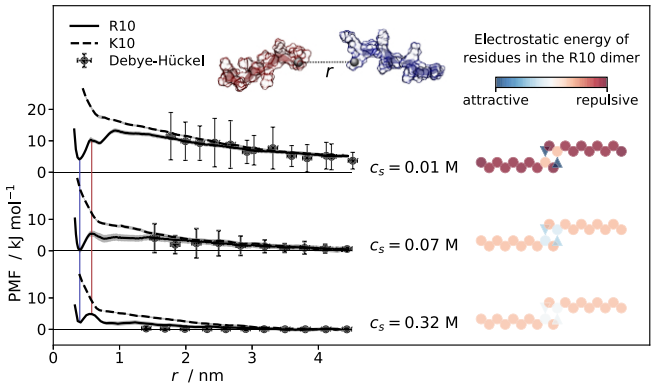

Fig. 2. (*Left*) PMFs calculated from umbrella-sampling MD simulations for pairs of R10 (solid line) and K10 (dashed line) molecules at $c_s = 0.01$ M, $c_s = 0.07$ M, and $c_s = 0.32$ M as a function of the separation between the guanidino-C and the ϵ-C atoms of the ninth residues. Shaded areas along the PMFs represent standard deviations (SD) of bootstrapped free energy profiles. Colored vertical lines connect maxima and minima that are common to all PMFs. Points represent free energy values calculated using the Debye–Hückel approximation, while error bars reflect SD of mass–center separations between peptides in the umbrella-sampling simulation windows. (*Top Center*) Generic representation of the deca-peptides where the ninth residues' guanidino-C/ϵ-C atoms are the black spheres and the dashed line between them represents the reaction coordinate, *r*. (*Right*) Schematic illustration of R10 dimeric structures. The circles and triangles represent positively charged arginine residues and negatively charged C-terminal carboxyl groups, respectively; while the coloring is based on the Debye–Hückel free energy calculated for each charge site interacting with all of the others, in the geometry adopted for the illustration.

moiety of the arginine side chains and are likely to be bidentate; whereas, in the K10 dimer, monodentate H bonds are formed between the C-terminal COO^- and the amino group of the lysine side chains.

^{1}H-^{13}C HSQC NMR Spectra. Fig. 4 shows the regions of 2D ^{1}H-^{13}C HSQC NMR spectra of R10 and R8KR corresponding to correlations between Cα and Hα atoms. Peptide solutions have $c_s = 0.025$ M, pH 5, and are measured at 293 K (*SI Results* and Fig. S4 for corresponding SAXS curves). With increasing c_p, R8KR spectra superpose almost perfectly, indicating the absence of significant peptide–peptide attractions for R8KR. For R10 small changes in chemical shifts are observed for both N-terminal and C-terminal residues. As $c_p/I(0)$ is proportional to the derivative of the osmotic pressure, Π, with respect to c_p, Fig. 4 *C* and *D* help to interpret the NMR data. At the lowest c_p, R10 solutions have significantly lower $(\partial\Pi/\partial c_p)_T$ at $c_s = 0.025$ M than at $c_s = 0.300$ M, while dilute R8KR solutions display a less pronounced and opposite trend for $(\partial\Pi/\partial c_p)_T$ vs. c_s. This suggests that at $c_s = 0.025$ M dimers may be present in the dilute $c_p = 1$-mM R10 solution, which also displays a different chemical shift for the N-terminal residue compared with the R8KR solution of $c_p = 5$ mM. The weak dependence of $c_p/I(0)$ on c_p observed for R10 at $c_s = 0.025$ M might be due to the repulsive nature of monomer–dimer and dimer–dimer interactions, as well as to a slight increase in dimer population. The latter is also reflected in the small changes in chemical shifts observed for R10 in the explored c_p range (Fig. 4*A*).

Occurrence of the Observed Mode of Interaction in Proteins. We searched a selection of 10,388 entries of the PDB (*SI Materials and Methods*) for the observed mode of interaction in R10 dimers, consisting of H bonding of two C-terminal COO^- groups with a Gdm^+–Gdm^+ ion pair. It was found to be present in 231 of the 1,697 protein crystal structures featuring at least one pair of stacked arginine residues (complete list in *SI Results*). The analysis focused on the COO^- groups of aspartate and glutamate side chains. However, in two proteins (PDB entries 5INJ and 4AZS), the Gdm^+–Gdm^+ pair is in the active site and interacts with the COO^- of the substrate. These enzymes are prenyltransferase PriB (19) and bifunctional methyl-transferase/kinase WbdD (20), while the substrates are tryptophan and *S*-adenosylmethionine, respectively.

Fig. 5 shows three characteristic structures which exemplify the geometries of the interacting Gdm^+ and COO^- moieties found in the crystal structures. In the most frequent arrangement I, each COO^- is H bonded with a different Gdm^+ moiety. In arrangement II, only one Gdm^+ moiety is involved in the interaction with two COO^- groups. Finally, in the least-occurring arrangement III, one of the COO^- groups is H bonded with one Gdm^+, whereas the second COO^- is approximately perpendicular to the molecular planes of the two Gdm^+ and forms H bonds with both Gdm^+ moieties in the ion pair. The residues involved in the interactions belong to different protein chains in 45.4% of the occurrences, while in the remaining 54.6% the interaction is between residues of the same chain. The analysis further showed that geometrical arrangements involving like-charge ion pairing of lysine side chains, stabilized by salt bridges with COO^- groups of aspartate and glutamate, are considerably less frequent than the mode of interaction characterized by arginine stacking (*SI Results*).

Discussion

In the SAXS curves of K10 at low c_s (dashed lines in Fig. 1) we observe a typical polyelectrolyte behavior—the intermolecular interactions are repulsive, causing $I(q)/c_p$ in the low-q range to decrease with increasing c_p, owing to the drop in osmotic compressibility (17). The high-q region is dominated by intramolecular scattering and is thus c_p independent. The maxima of $I(q)/c_p$ observed at $q = q_{max}$ result from intermolecular distance correlations (18); the inverse of q_{max} is related to the mean distance between nearest neighbors in solution. SAXS curves collected for R10 samples at $c_s = 0.060$ M and $c_s = 0.150$ M (solid lines in Fig. 1 *A* and *B*) reveal a striking difference in intermolecular interactions compared with the case of K10. The increase of $I(q)/c_p$ in the low-q range with increasing c_p corresponds to an increase in osmotic compressibility, indicating a net attractive interaction between the R10 peptides. That is, in R10 solutions at intermediate c_s an attraction dominates over the repulsive double-layer force.

For solutions of $c_s = 0.300$ M (Fig. 1*C*), where electrostatic interactions are effectively screened, scattering intensities for K10 at increasing c_p suggest an excluded-volume effect, i.e., an

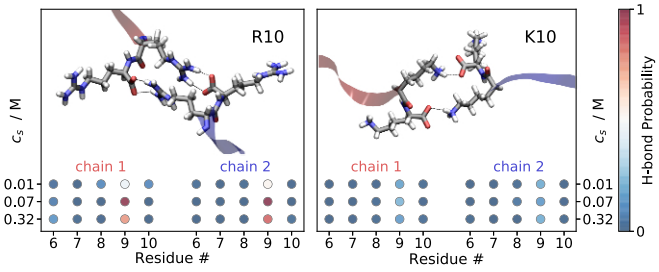

Fig. 3. Snapshots from MD simulations illustrating the interaction between C-terminal residues of two R10 and two K10 molecules at close separation. Dashed lines represent the hydrogen bonds between the negatively charged carboxyl groups and the positively charged guanidinium and ammonium moieties of the ninth residues. Colored circles represent the probability of H bonds between the carboxyl groups of one peptide and the last five residues of the other peptide. Snapshots and probabilities are obtained from MD simulations with separations of 0.43 ± 0.04 nm and 0.74 ± 0.06 nm between the ninth residues' guanidino-C and ϵ-C atoms, respectively.

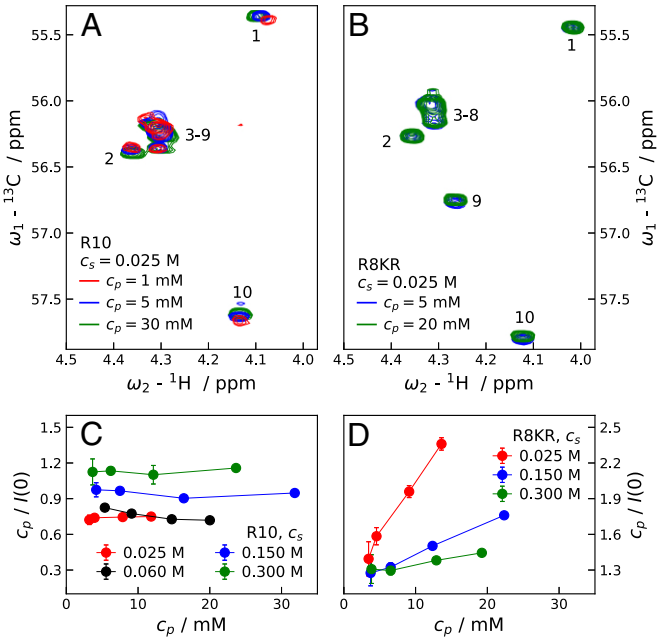

Fig. 4. (*A* and *B*) Regions corresponding to the correlations between Cα and Hα atoms of 2D ^{1}H-^{13}C HSQC NMR spectra of R10 (*A*) and R8KR (*B*) at various peptide concentrations, c_p, pH 5, and ionic strength 0.025 M. Cross-peaks are labeled with the residue numbers to which they are assigned. (*C* and *D*) Extrapolated $c_p/I(0)$ values for samples of R10 (*C*) and R8KR (*D*) of increasing c_p and c_s at pH 7.8. Error bars represent SD. Lines connecting points in *C* and *D* are guides to the eye.

effective repulsion caused by the loss of entropy in the more crowded solutions. For R10 at $c_s = 0.300$ M (Fig. 1*C*), the effect of c_p on the scattering intensities shows a nonmonotonic trend in $I(q)/c_p$ at small angles, indicating that the attractive interaction competes with the repulsive excluded-volume effect. Comparing SAXS curves for R10 at low and high c_s, it can be inferred that the attractive interaction has an important electrostatic component. The electrostatic adhesive force is strong at low c_s, where it overpowers the electrostatic repulsion between the highly charged peptides. In contrast, a much weaker attraction is required to compensate for the excluded-volume effect. In Fig. 1*D*, the decreasing trends of $I(0)/c_p$ values for K10, K8RK, and R8KR solutions show that the net interaction between these molecules is repulsive, whereas the large and increasing $I(0)/c_p$ values for R10 solutions of increasing c_p indicate that at c_s = 0.150 M a fraction of R10 molecules self-associates. SAXS data provide evidence that the ninth arginine residue is essential for the attractive interaction; nonetheless, as shown by the scattering curves for K8RK molecules (*SI Results* and Fig. S2), the remaining arginine residues have a role in lowering the electrostatic repulsion.

The PMFs (Fig. 2) show that K10 molecules repel each other at all length scales and salt concentrations, while R10 molecules, despite bearing the same net charge, display a salt concentration-dependent attraction at short separation. This result is qualitatively insensitive to the choice of the force field (*SI Results* and Figs. S5–S7). A minimum, at 0.4 nm separation, and an adjacent maximum, at 0.58 nm separation, are present in the PMFs of R10 at all c_s, indicated by the blue and red vertical lines in Fig. 2. The free energy difference between the maximum and the minimum serves as an estimate of the strength of the adhesive force. The fact that the free energy difference diminishes with increasing c_s confirms that the adhesive force has an important electrostatic component. The stability of the dimeric species is indicated by the free energy values of the minima in the PMFs of R10 and depends on the interplay between the double layer and the electrostatic adhesive force. From our simulation results, it can be inferred that the balance between the two forces, which is modulated by the ionic strength, favors the attraction at the intermediate $c_s = 0.07$ M, where the R10 dimer corresponds to the lowest free energy value. This result can be explained by a simple electrostatic argument, schematically illustrated in Fig. 2. The color scheme exemplifies that with increasing c_s the repulsion between the outer positively charged sites is screened more efficiently than the attraction of oppositely charged sites in the binding region between the two peptides. As a result of the different efficiency by which the short-range attraction and the long-range repulsion are screened with increasing c_s, the net Debye–Hückel free energy of the dimer shows a nonmonotonic trend—the screening majorly affects like-charged side chains that are farther apart until c_s is so large that even the short-range attractions in the binding region are effectively screened.

Competition between double-layer and electrostatic adhesive forces has been observed for other biomolecules, e.g., a monoclonal antibody mAb1 (1) and a globular milk protein lactoferrin (2). In both cases, the attraction stems from a charged patch on the protein surface and becomes dominant at a critical salt concentration.

MD simulations provide a detailed picture of the origin of the attraction between R10 peptides. The last two terminal residues of R10 can be identified as an adhesive patch displaying two oppositely charge sites. The positive site is the Gdm^+ moiety of the ninth arginine residue while the negative site is the C-terminal carboxyl group (COO^-). We suggest that the observed attraction between R10 molecules occurs through complementarity of the charged groups in the adhesive patch of two interacting peptides. When two patches are at close separation, the Gdm^+ moieties of the ninth residues of the peptides form an ion pair. The divalent charge site, generated by the stacking of the arginine side chains, is stabilized by two intermolecular hydrogen bonds. These are salt bridges formed between the negatively charged C-terminal COO^- and the positively charged Gdm^+ of the ninth residues. Theoretical and experimental studies suggest that Gdm^+ ions form weakly stable like-charge ion pairs in aqueous medium (21–28). The free energy of interaction between two Gdm^+ ions is minimized when they are stacked parallel to each other in staggered geometry with carbon atoms with separation between 0.35 nm and 0.46 nm (28). The weak Gdm^+–Gdm^+ attraction in water is estimated to be around -2 kJ·mol^{-1} and is due to a combination of quadrupole–quadrupole interaction, dispersive forces, and the hydrophobic effect (28). The hydrophobicity of the surfaces of Gdm^+, in conjunction with the ability of Gdm^+ to form H bonds with functional groups in its molecular plane, has recently been confirmed from the analysis of interactions of arginine side chains in proteins in the PDB (29).

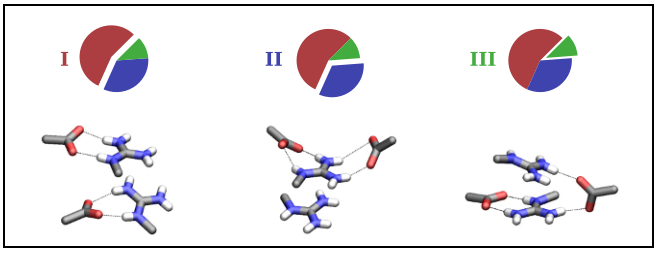

Fig. 5. Summary illustration of the analysis of the PDB of the interaction between the Gdm^+ ion pair and the two carboxyl groups. The 231 selected protein structures are divided into three categories differing in the geometrical arrangement of the groups involved in the interaction. From *Bottom* to *Top*: Structures exemplifying the geometrical arrangements and pie charts showing the incidence of the arrangements.

Our analysis of H bonding involving C termini of R10 or K10 (Fig. 3) shows that the probability of forming H bonds between the ninth residues at close separation and COO^- is remarkably higher in R10 than in K10. Since salt bridges involving COO^- and ammonium (NH_3^+) or Gdm^+ moieties have comparable strengths (30), the higher probability observed for R10 highlights the important role of the arginine side chain in the R10–R10 patchy attraction. Due to the approximately tetrahedral geometry of NH_3^+, restraining the position of two lysine side chains at close separation hinders the formation of H bonds with COO^-. In contrast, the H bonding of Gdm^+ occurs in the molecular plane and is unaffected by stacking. The probability for the formation of salt bridges involving the ninth residues of R10 and K10 has a nonmonotonic dependence on c_s. The formation of salt bridges is the consequence of short-range attractive electrostatic interactions. Therefore, the trend can be rationalized by the same argument used to explain the c_s dependence of the free energy minima in the PMFs of R10.

Our results highlight the fact that an important component of the described patchy attraction is of electrostatic origin and is imparted by two salt bridges. The main difference in the interaction between R10 and K10 lies in the favorable Gdm^+–Gdm^+ stacking. As shown by the H-bond probabilities calculated for K10, in the absence of pairing of the positively charged side chains, C-terminal salt bridges are substantially less energetically favorable.

The NMR data confirm that R8KR solutions are monodisperse in the explored c_p range, while R10 molecules self-associate. The observed changes in chemical shifts in Fig. 4*A* are consistent with a nonspecific peptide–peptide interaction characterized by the stacking of side chains of several arginine residues of the R10 peptide. Accordingly, MD simulations indicate that, besides the 9th residue, the 8th and 10th residues also contribute to the attractive interaction (*SI Results* and Fig. S8) and that the stacking of Gdm^+ moieties occurring between the remaining residues can be stabilized by chloride ions (31) (*SI Results* and Fig. S9). For the N-terminal residues, the changes in chemical shift with increasing c_p may be explainable by the favorable interaction of the N terminus of one peptide with the binding region of the R10 dimer. Indeed, when the side chains of the 10th and 8th residues are constrained at stacking separations, the PMF as a function of distance between guanidino-C atoms of 10th and 2nd residues of R10 shows two local minima at close separations (*SI Results* and Fig. S10).

The mode of interaction responsible for the adhesive force between R10 molecules is well represented in the PDB (Fig. 5). A total of 231 X-ray crystal structures of 10,388 entries present at least one ion pair formed by the stacking of two arginine side chains, which interacts with two carboxyl groups of aspartate or glutamate residues. Arginine residues are also found at protein–protein interfaces more frequently than lysine residues, reflected in the "stickiness" scale proposed by Levy and coworkers (32).

The concentration effect observed in the SAXS data from the dilution series of R10 at low-to-intermediate c_s (Fig. 1, *SI Results*, and Fig. S2 *A* and *C*) is consistent with a monomer–oligomer equilibrium where the population of oligomeric forms in solution increases with increasing c_p. This is further supported by concentration-dependent NMR ^1H and ^{13}C chemical shifts (Fig. 4), assuming fast exchange on the NMR chemical shift timescale between monomeric and oligomeric forms.

Self-aggregation has been recently shown to be important for the effective translocation of a 9-aa-long peptide across the plasma membrane (33). As a consequence, the propensity of R10 molecules to aggregate may contribute to the explanation of the high cellular uptake of RRPs.

The transduction efficiency of oligo-arginines depends on the number of residues. It is maximal for chains of 6–15 aa, but considerably lower for shorter as well as longer peptides (7, 12). As the number of positively charged residues increases, the force balance governing the self-association of oligo-arginines gradually moves toward the repulsive double-layer force. The presence of an upper limit in the range of optimal chain lengths may reflect the importance of self-association in the transduction of oligo-arginines.

Conclusions

We have used a combination of MD simulations with SAXS and NMR experiments with the aim of elucidating the molecular mechanism of self-association of arginine-rich oligopeptides. Concentration studies point to an important electrostatic component of the attraction between R10 molecules, while single-point mutations underline a binding motif involving the C terminus and the adjacent arginine residues. The present results not only support the notion of the important role of self-aggregation in the transduction of cell-penetrating peptides, but also open the path to future studies of the potential biological roles of the newly discovered binding motif, as exemplified by its abundant occurrence in the PDB.

ACKNOWLEDGMENTS. We are grateful to Dr. Petra Pernot and Dr. Martha Brennich at the European Synchrotron Radiation Facility (ESRF), Grenoble, for providing assistance in using beamline BM29 and to Ellen Rieloff for experimental assistance. We thank the Swedish Research Council, the Swedish Foundation for Strategic Research, Lunarc in Lund for computational resources, and ESRF for providing beam time. This study was supported by the Science Faculty project grant program for research with neutrons and synchrotron light [Lund University (LU) Strategic funds for MAX-IV and European Spallation Source (ESS); Grant V2016/1301]. P.J. acknowledges the Czech Science Foundation for support via Grant 16-01074S. This work used the platforms of the Grenoble Instruct-ERIC Center [Integrated Structural Biology Grenoble (ISBG): Unité Mixte de Service (UMS) 3518 CNRS-Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA)-Univ. Grenoble Alpes (UGA)-European Molecular Biology Laboratory (EMBL)] with support from French Infrastructure for Integrated Structural Biology (FRISBI) (ANR-10-INSB-05-02) and Grenoble Alliance for Integrated Structural Cell Biology (GRAL) (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology.

1. Roberts D, et al. (2014) The role of electrostatics in protein–protein interactions of a monoclonal antibody. *Mol Pharm* 11:2475–2489.
2. Li W, et al. (2015) Charge-induced patchy attractions between proteins. *J Phys Chem B* 119:503–508.
3. Li W, Persson BA, Lund M, Bergenholtz J, Oskolkova MZ (2016) Concentration-induced association in a protein system caused by a highly directional patch attraction. *J Phys Chem B* 120:8953–8959.
4. Ohnuki J, Yodogawa A, Takano M (2017) Electrostatic balance between global repulsion and local attraction in reentrant polymerization of actin. *Cytoskeleton*, 10.1002/cm.21391.
5. Roberts CJ, Blanco MA (2014) Role of anisotropic interactions for proteins and patchy nanoparticles. *J Phys Chem B* 118:12599–12611.
6. Dumetz AC, Snellinger-O'Brien AM, Kaler EW, Lenhoff AM (2007) Patterns of protein-protein interactions in salt solutions and implications for protein crystallization. *Protein Sci* 16:1867–1877.
7. Mitchell D, Steinman L, Kim D, Fathman C, Rothbard J (2000) Polyarginine enters cells more efficiently than other polycationic homopolymers. *J Pept Res* 56:318–325.
8. Strömstedt AA, Pasupuleti M, Schmidtchen A, Malmsten M (2009) Oligotryptophan-tagged antimicrobial peptides and the role of the cationic sequence. *Biochim Biophys Acta* 1788:1916–1923.
9. Vazdar M, et al. (2013) Aggregation of oligoarginines at phospholipid membranes: Molecular dynamics simulations, time-dependent fluorescence shift, and biomimetic colorimetric assays. *J Phys Chem B* 117:11530–11540.
10. Wu Z, Cui Q, Yethiraj A (2013) Why do arginine and lysine organize lipids differently? Insights from coarse-grained and atomistic simulations. *J Phys Chem B* 117:12145–12156.
11. Li L, Vorobyov I, Allen TW (2013) The different interactions of lysine and arginine side chains with lipid membranes. *J Phys Chem B* 117:11906–11920.
12. Futaki S, et al. (2000) Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. *J Biol Chem* 276:5836–5840.
13. Tünnemann G, et al. (2008) Live-cell analysis of cell penetration ability and toxicity of oligo-arginines. *J Pept Sci* 14:469–476.

14. Lättig-Tünnemann G, et al. (2011) Backbone rigidity and static presentation of guanidinium groups increases cellular uptake of arginine-rich cell-penetrating peptides. *Nat Commun* 2:453.
15. Sun D, Forsman J, Lund M, Woodward CE (2014) Effect of arginine-rich cell penetrating peptides on membrane pore formation and life-times: A molecular simulation study. *Phys Chem Chem Phys* 16:20785–20795.
16. Robison AD, et al. (2016) Polyarginine interacts more strongly and cooperatively than polylysine with phospholipid bilayers. *J Phys Chem B* 120:9287–9296.
17. Nierlich M, et al. (1979) Small angle neutron scattering by semi-dilute solutions of polyelectrolyte. *J Phys* 40:701–704.
18. Hayter J, Janninck G, Brochard-Wyart F, de Gennes P (1980) Correlations and dynamics of polyelectrolyte solutions. *J Phys Lett* 41:451–454.
19. Elshahawi SI, et al. (2017) Structure and specificity of a permissive bacterial c-prenyltransferase. *Nat Chem Biol* 13:366–368.
20. Hagelueken G, et al. (2012) Structure of WbdD: A bifunctional kinase and methyltransferase that regulates the chain length of the O antigen in Escherichia coli O9a. *Mol Microbiol* 86:730–742.
21. Mason PE, Neilson GW, Dempsey CE, Barnes AC, Cruickshank JM (2003) The hydration structure of guanidinium and thiocyanate ions: Implications for protein stability in aqueous solution. *Proc Natl Acad Sci USA* 100:4557–4561.
22. Mason PE, et al. (2004) The structure of aqueous guanidinium chloride solutions. *J Am Chem Soc* 126:11462–11470.
23. Mason PE, Brady JW, Neilson GW, Dempsey CE (2007) The interaction of guanidinium ions with a model peptide. *Biophys J* 93:L04–L06.
24. Vazdar M, Vymětal J, Heyda J, Vondrášek J, Jungwirth P (2011) Like-charge guanidinium pairing from molecular dynamics and ab initio calculations. *J Phys Chem A* 115:11193–11201.
25. Vazdar M, Uhlig F, Jungwirth P (2012) Like-charge ion pairing in water: An ab initio molecular dynamics study of aqueous guanidinium cations. *J Phys Chem Lett* 3:2021–2024.
26. Wernersson E, et al. (2012) Counterion condensation in short cationic peptides: Limiting mobilities beyond the Onsager-Fuoss theory. *Electrophoresis* 33:981–989.
27. Shih O, et al. (2013) Cation-cation contact pairing in water: Guanidinium. *J Chem Phys* 139:035104.
28. Allolio C, Baxova K, Vazdar M, Jungwirth P (2016) Guanidinium pairing facilitates membrane translocation. *J Phys Chem B* 120:143–153.
29. Armstrong CT, Mason PE, Anderson JLR, Dempsey CE (2016) Arginine side chain interactions and the role of arginine as a gating charge carrier in voltage sensitive ion channels. *Sci Rep* 6:21759.
30. Debiec KT, Gronenborn AM, Chong LT (2014) Evaluating the strength of salt bridges: A comparison of current biomolecular force fields. *J Phys Chem B* 118:6561–6569.
31. Formaneck MS, Ma L, Cui Q (2006) Effects of temperature and salt concentration on the structural stability of human lymphotactin: Insights from molecular simulations. *J Am Chem Soc* 128:9506–9517.
32. Levy ED, De S, Teichmann SA (2012) Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc Natl Acad Sci USA* 109:20461–20466.
33. Macchi S, et al. (2015) Spontaneous membrane-translocating peptides: Influence of peptide self-aggregation and cargo polarity. *Sci Rep* 5:16914.
34. Best RB, Zheng W, Mittal J (2014) Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association. *J Chem Theory Comput* 10:5113–5124.
35. Henriques J, Cragnell C, Skepö M (2015) Molecular dynamics simulations of intrinsically disordered proteins: Force field evaluation and comparison with experiment. *J Chem Theory Comput* 11:3420–3431.
36. Minh DDL (2007) Multidimensional potentials of mean force from biased experiments along a single coordinate. *J Phys Chem B* 111:4137–4140.
37. Jong DHD, et al. (2011) Determining equilibrium constants for dimerization reactions from molecular dynamics simulations. *J Comput Chem* 32:1919–1928.
38. Hicks M, Gebicki JM (1986) Rate constants for reaction of hydroxyl radicals with tris, tricine and hepes buffers. *FEBS Lett* 199:92–94.
39. Grishaev A (2001) Sample preparation, data collection, and preliminary data analysis in biomolecular solution X-Ray scattering. *Curr Protoc Protein Sci* 70:17.14.1–17.14.18.
40. Kuipers BJH, Gruppen H (2007) Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis. *J Agric Food Chem* 55:5445–5451.
41. Pernot P, et al. (2013) Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J Synchrotron Radiat* 20:660–664.
42. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) Primus: A windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* 36:1277–1282.
43. Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 74:829–836.
44. Jacques DA, Trewhella J (2010) Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. *Protein Sci* 19:642–657.
45. Kharakoz DP (1997) Partial volumes and compressibilities of extended polypeptide chains in aqueous solution: Additivity scheme and implication of protein unfolding at normal and high pressure. *Biochemistry* 36:10276–10285.
46. Whitten AE, Cai S, Trewhella J (2008) MULCh: Modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies. *J Appl Crystallogr* 41:222–226.
47. Torrie G, Valleau J (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comput Phys* 23:187–199.
48. Pronk S, et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29:845–854.
49. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126:014101.
50. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52:7182–7190.
51. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An n·log(n) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092.
52. Hess B (2008) P-LINCS: A parallel linear constraint solver for molecular simulation. *J Chem Theor Comput* 4:116–122.
53. Hub JS, de Groot BL, van der Spoel D (2010) g_wham—a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J Chem Theor Comput* 6:3713–3720.
54. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graphics* 14:33–38.
55. Wernet P (2004) The structure of the first coordination shell in liquid water. *Science* 304:995–999.
56. McGibbon RT, et al. (2015) MDTraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophys J* 109:1528–1532.
57. Wang G, Dunbrack RL (2003) PISCES: A protein sequence culling server. *Bioinformatics* 19:1589–1591.
58. Berman HM (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242.
59. Eastman P, et al. (2013) OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *J Chem Theor Comput* 9: 461–469.