

Coarse-grained modelling of the intrinsically disordered protein Histatin 5 in solution. Monte Carlo simulations in combination with SAXS.

Carolina Cragnell,^{*,†} Dominique Durand,[‡] Bernard Cabane,[¶] and Marie Skepö[†]

Chemical department, Theoretical Chemistry, Lund University, Lund, Sweden, I2BC, Université Paris-Sud, ORSAY Cédex, France, and PMMH, CNRS UMR 7636, ESPCI, F-75231 Paris cedex 05, France

E-mail: carolina.cragnell@teokem.lu.se

KEYWORDS: Antimicrobial activity, Monte Carlo simulations, Unstructured proteins, Cationic protein, Small Angle X-ray Scattering

*To whom correspondence should be addressed

[†]Chemical department, Theoretical Chemistry, Lund University, Lund, Sweden

[‡]I2BC, Université Paris-Sud, ORSAY Cédex, France

[¶]PMMH, CNRS UMR 7636, ESPCI, F-75231 Paris cedex 05, France

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/prot.25025

© 2016 Wiley Periodicals, Inc.

Received: Nov 10, 2015; Revised: Feb 17, 2016; Accepted: Feb 18, 2016

Abstract

Monte Carlo simulations and coarse-grained modelling have been used to analyze Histatin 5, which is an unstructured short cationic salivary peptide known to have anti-candidal properties. The calculated scattering functions have been compared with intensity curves and the distance distribution function $P(r)$ obtained from SAXS, at both high and low salt concentrations. The aim is to achieve a molecular understanding and a physico-chemical insight of the obtained SAXS results and to gain information of conformational changes of Histatin 5 due to altering salt content, charge distribution, and net charge. From a modelling perspective, the accuracy of the electrostatic interaction is of special interest. The used coarse-grained model is based on the primitive model in which charged hard spheres differing in charge and in size represent the ionic particles, and the solvent only enters the model through its relative permittivity. The Hamiltonian of the model comprises three different contributions: (i) excluded volumes, (ii) electrostatic, and (iii) van der Waals interactions. Even though the model can be considered as gross omitting atomistic details, a great correspondence is obtained with experimental results.

Introduction

Saliva proteins are charged macromolecules that are of great importance from a health perspective. When humans are under heavy medication, for example cell toxins and radiation therapy, or are prescribed other medicines or medical treatments; an unpleasant side effect is saliva malproduction, both with respect to composition and amount. Saliva contains approximately 99.5 % water, 0.2 % proteins, as well as salts and cellular components. Even though the amount of proteins seems very low, almost negligible (!), they play an important role for the oral health; and many of the proteins possess multifunctionality. For example, the proteins are important for remineralization of the enamel, lubrication of the teeth and gums, and to prevent bacteria and viruses from damaging the body. The proteins are also

essential for the ability of saliva to form oral films on solids. This film i.e. the pellicle, is important for the maintenance of oral health and surface integrity.¹

Several of the saliva proteins, such as the proline rich proteins, the statherins, and the histatins, belong to the group of intrinsically disordered proteins (IDPs).² These proteins are characterized by lack of stable tertiary structure under physiological conditions in vitro. More recently, it has been shown that approximately 30 % of all proteins in eukaryotic organisms belong to this group and that IDPs are involved in a large number of central biological processes and diseases.³ This discovery challenges the traditional protein structure paradigm, which states that a specific well-defined structure is required for the correct function of a protein. Biochemical evidence has since shown that IDPs are functional, and that the lack of folded structures is related to their functions.^{4,5}

In this study Histatin 5 (His5) has been used as model protein. His5 is a short multifunctional cationic saliva peptide with a molecular weight of approximately 3 kDa (24 amino acids). His5 belongs to the Histatin family of proteins, and they act as the first line of defence against oral candidiasis caused by *Candida albicans*.^{6,7} Furthermore, they also possess bactericidal effects and bind polyphenolic compounds as tannin.^{8,9} The Histatin family consists of 12 members where His5 is the most potent with respect to its antifungal activity. The antimicrobial activity and the interaction with tannins have been ascribed to the high content of basic amino acids. His5 also participate in the formation of a protective layer (pellicle) on smooth tooth surfaces,¹⁰ and thereby prevent microbial colonization and stabilize mineral-solute interactions. From NMR and CD-measurements it is known that His5 has a flexible structure at physiological conditions and the protein is considered to belong to the IDPs.^{11,12} Full atomistic molecular dynamic simulations of His5 have been performed recently by our group. Our simulations produced results in very good agreement with the experimentally obtained Kratky plot, and intrapeptide distances obtained through the pair distance distribution function, $P(r)$. For the interested reader please find the paper by Henriques et al¹³ and the references therein.

In this study we have used Monte Carlo simulations and coarse-grained modelling in combination with small angle x-ray scattering (SAXS) to analyse His5 SAXS-spectra at both high and low salt concentrations as well as varying peptide concentrations. The peptide model is based on the primitive model¹⁴ in which hard spheres differing in charge and in size represent the ionic particles, and the solvent only enters the model through its relative permittivity. The Hamiltonian of model comprises three different contributions: (i) excluded volumes, (ii) electrostatic interactions, and (iii) van der Waals interactions. Even though the model can be considered as gross omitting all atomistic details, a great correspondence is obtained when comparing simulated and experimental intensity curves.

The aim is to find an answer the following question: Is it possible to use a coarse-grained model based on the primitive model to understand the physico-chemical properties of His5 and to capture the effect of electrostatic interactions? Since His5 has a relatively small size, and since it is monomeric, makes it a perfect model protein for both atomistic^{15,16} and coarse-grained modelling.¹⁷ These results are also important for understanding how the lack of structure in solution relates to the anti-candidacidal properties of the peptide when it is interacting with a surface, for example the cell membrane.^{18,19}

Modelling studies of IDPs of both saliva^{17,20,21} and milk proteins^{21,22} have been performed by our group before, although here we make a direct comparison with experimental results.

Materials and Methods

SAXS

The SAXS measurements were carried out using BM29 beamline at the European synchrotron facility ESRF in Grenoble, France. The incident beam wavelength was 0.99 Å and the distance between the sample and the PILATUS 1M detector was set to 2867 mm giving a scattering vector range of $0.0028 \text{ \AA}^{-1} < q > 0.45 \text{ \AA}^{-1}$. The scattering vector is defined as $q = 4\pi \sin\theta/\lambda$ where 2θ is the scattering angle. For each sample and pure solvent, several

successive frames (typically 10 - 25) of 1 s, were recorded and analyzed. Special attention was paid to radiation damage by comparing the successive frames prior further processing of the data. The average scattering from each individual frame was computed as well as the experimental error. The background, i.e. pure solvent, was subtracted from the corresponding sample solution. All measurements were performed at 20 °C. $I(0)$ was converted to absolute scale by measuring the scattering of water. A few additional measurements were performed on beamline SWING at the Synchrotron SOLEIL. Distance to detector (Aviex CCD) 1501 mm giving a scattering vector range of $0.008 \text{ \AA}^{-1} < q > 0.73 \text{ \AA}^{-1}$.

Histatin 5 was purchased from American Peptide Company, CAS: [72-2-25]. The counterion was trifluoroacetate and the appearance white lyophilized powder. The Tris (Saveen Werner AB with a purity of > 99.9%, CAS: [77-86-1]) 10 mM buffer was prepared in Milli-Q water and acidified to pH 7 with HCl. The ionic strength (IS) of the solutions was adjusted with NaCl (Scharlau with a purity of > 99.5% CAS: [7647-14-5]). Prior to adding the peptide to the buffer, the buffer was filtered through a $0.2 \mu\text{m}$ hydrophilic polypropylene membrane (Pall Corporation). After mixing the protein powder with the buffer, a concentration cell (Vivaspin 20, MWCO 1 kDa, Prod. No. VS2002, Sartorius Stedim Biotech GmbH, Germany), was used to remove low molecular weight impurities including divalent ions from the freeze-dried sample. The sample was rinsed with buffer corresponding to at least 20 times the sample volume. For this purpose, centrifugations at 3500 rpm at 18 °C were applied. In addition, the samples were dialyzed for 12 hours to ensure exact background match (Slide-A-Lyzer MINI, MWCO 2 kDa, Prod. No. 69580, Thermo Scientific, U.S.A.). Before the SAXS measurements the samples were centrifuged with an ultracentrifuge (rotor TLA55) at 13000 rpm for at least 30 min to remove aggregates etc. The protein concentrations were measured after preparation and then immediately before the experiment using a Nanodrop spectrophotometer at the beamline, $\varepsilon = 2560 \text{ M}^{-1}\text{cm}^{-1}$, $\lambda = 280 \text{ nm}$. The studied His5 concentrations were varied in order to overlap the physiological protein concentration range in saliva (1 - 3.5 mg/mL). In addition, higher concentrations were studied to obtain a more

complete understanding of the interaction between the peptides. The studied concentration range was set to 1.1 - 7.4 mg/mL. The ionic strength (IS) was varied from 10 mM to 150 mM by adding NaCl.

Nano LC MS/MS

Mass spectrometry was employed to ensure the purity of the His5 samples. The experiment was performed with an EasyLC nanoflow HPLC interfaced with a nanoEasy spray ion source (Proxeon Biosystems, Odense, Denmark) connected to a Fusion Orbitrap mass spectrometer (Thermo Fisher Scientific). The sample was loaded on a PepMap column 2-cm (75- μ m inner diameter packed with 3 μ m resin) and the chromatographic separation was performed at 35 °C on a 25-cm (75- μ m inner diameter) EASY-Spray column packed with 2 μ m of resin (Proxeon Biosystems). The nanoHPLC was operating at 300 nL/min with a gradient of 5 - 22% solvent B (0.1% (v/v) formic acid, 100% (v/v) acetonitrile in water) in solvent A (0.1% (v/v) formic acid in water) during 20 min and then 22 - 32% during 2 min followed with an increase to 98% B during 2 min. A MS scan (350 - 1500 m/z) was recorded in the Orbitrap mass analyzer set at a resolution of 60,000 at 400 m/z , 1 x 10 automatic gain control target, and 500-ms maximum ion injection time. The mass spectrometric conditions were as follows: spray voltage, 1.9 kV; no sheath or auxiliary gas flow; S-lens 60%; ion transfer tube temperature, 275 °C. Figures from the measurements are shown in supplementary.

Simulations

Model

The residues (24 amino acids, aa) of His5 are represented by hard spheres, mimicking the residue excluded volume, and are connected via harmonic bonds. The N- and C-terminal are included explicitly giving a total chain of 26 beads. The initial bead radius was set to 2 Å providing a realistic contact separation between the charges and an accurate Coulomb

interaction. The nonbonded spheres interact through van der Waals and electrostatic interactions. The inter-particle electrostatics are described at the Debye-Hückel level. The simulations are carried out at constant pH with fixed point charges. Each residue is either negative, positive or neutral dependent on the MC simulated average protonation state of the amino acid, determined by the in-house program Faunus.²³ The total potential energy of the simulated system contains both bonded and nonbonded contributions and is given by:

$$U_{\text{tot}} = U_{\text{nonbond}} + U_{\text{bond}} = U_{\text{hs}} + U_{\text{el}} + U_{\text{short}} + U_{\text{bond}} \quad (1)$$

where the nonbonded energy is assumed pairwise additive according to:

$$U_{\text{nonbond}} = \sum_{i < j} u_{ij}(r_{ij}) \quad (2)$$

The excluded volume is taken into account through the hard-sphere potential, U_{hs} , given by:

$$U_{\text{hs}} = \sum_{i < j} u_{ij}^{\text{hs}}(r_{ij}) \quad (3)$$

which sums up over all amino acids. The hard-sphere potential, $u_{ij}^{\text{hs}}(r_{ij})$, between two beads in the model is given by:

$$u_{ij}^{\text{hs}}(r_{ij}) = \begin{cases} 0, & r_{ij} \geq R_i + R_j \\ \infty, & r_{ij} < R_i + R_j \end{cases} \quad (4)$$

where $r_{ij} \equiv |\mathbf{R}_i - \mathbf{R}_j|$ is the center to center distance between two beads. \mathbf{R} refers to the coordinate vector. In the Hamiltonian, the electrostatic energy potential, U_{el} , is given by an extended Debye-Hückel potential:

$$U_{\text{el}} = \sum_{i < j} u_{ij}^{\text{el}}(r_{ij}) = \sum_{i < j} \frac{Z_i Z_j e^2 \exp[-\kappa(r_{ij} - (R_i + R_j))]}{4\pi\epsilon_0\epsilon_r (1 + \kappa R_i)(1 + \kappa R_j)} \frac{1}{r_{ij}} \quad (5)$$

where $\kappa = [(e^2/\epsilon_0\epsilon_r kT) \sum_i Z_i^2 \rho_1]^{1/2}$ ϵ_r and ϵ_0 refers to the relative permittivity of water (78.4) at 298 K and vacuum respectively, which are assumed to be constant throughout the system. The van der Waals interaction (vdW) between the beads is included through an approximate arithmetic average over all amino acids, represented by a short range attraction given by:

$$U_{\text{short}} = - \sum_{i < j} \frac{\epsilon}{r_{ij}^6} \quad (6)$$

where ϵ reflects the polarizability of His5 and thus sets the strength of the interaction. ϵ was set to $0.6 \cdot 10^4$ kJ \AA^6 /mol giving an attractive potential of 0.6 kT at closest contact. The bonded energy is described by a harmonic potential given by:

$$U_{\text{bond}} = \sum_{i=1}^{N_{\text{seg}}-1} \frac{k_{\text{bond}}}{2} (r_{i,i+1} - r_0)^2 \quad (7)$$

where $r_{i,i+1}$ refers to the interbead distance and r_0 , 4.1 \AA , the equilibrium distance. The force constant k_{bond} is set to 0.4 N/m, and the protein is assumed to be flexible.

The fluctuating titrating charges of the amino acid beads are handled through the titrating energy given by:

$$\sum_{i=1}^{n_p} k_B T (\text{pH} - \text{p}K_{a,i}) \ln 10 \quad (8)$$

where n_p refers to number of protonated beads. The intrinsic $\text{p}K_a$ values were taken from Nozaki and Tanford.²⁴

Structural Analysis

The model was validated by comparing the simulated structure factors with the experimental scattering intensities obtained by SAXS. For a system containing N scattering objects the structure factor is given by:

$$S(q) = \left\langle \frac{1}{N} \left| \sum_{j=1}^N \exp(i\mathbf{q} \cdot \mathbf{r}_j) \right|^2 \right\rangle \quad (9)$$

The total structure factor can further be decomposed into partial structure factors given by:

$$S_{ij}(q) = \left\langle \frac{1}{(N_i N_j)^{1/2}} \left[\sum_{i=1}^{N_i} \exp(i\mathbf{q} \cdot \mathbf{r}_i) \right] \left[\sum_{j=1}^{N_j} \exp(-i\mathbf{q} \cdot \mathbf{r}_j) \right] \right\rangle \quad (10)$$

The total and partial $S(q)$ are related through:

$$S(q) = \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \frac{(N_i N_j)^{1/2}}{N} S_{ij}(q) \quad (11)$$

In order to account for an approximate effective particle/residue form factor, the scattering profile, $I(q) = S(q)S(q)^*$, further need an appropriate normalization, such that $I(q = 0)$ coincides with the experimental scattering profile. Note that the lower limit of q , the wave vector, is controlled by the box length in the simulations according to $q_{\text{low}} = 2\pi/L_{\text{box}}$. The conformational changes of His5 were quantified by the radius of gyration, $\langle R_g^2 \rangle^{1/2}$, and end-to-end distance, $\langle R_{ee}^2 \rangle^{1/2}$:

$$\langle R_g^2 \rangle^{1/2} \equiv \langle N^{-1} \sum_{i=1}^N (\mathbf{r}_i - \mathbf{r}_{\text{com}})^2 \rangle^{1/2} \quad (12)$$

$$\langle R_{ee}^2 \rangle^{1/2} \equiv \langle |\mathbf{r}_1 - \mathbf{r}_N|^2 \rangle^{1/2} \quad (13)$$

where N refers to the number of beads in the chain, $\langle \cdot \cdot \rangle$ to an ensemble average, and \mathbf{r}_{com} is the center of mass.

Simulation Aspects

The equilibrium properties of the model systems were obtained applying Monte Carlo simulations (MC) carried out in the canonical ensemble, i.e. constant volume, number of beads, and temperature ($T = 298$ K) utilizing the Metropolis algorithm.²⁵ 40 chains were enclosed in a cubic box with altering volume, which was concentration dependent (typical box length ~ 400 Å). Periodic boundary conditions were applied in all directions. The long-ranged

Coulomb interactions were truncated using the minimum image convention. Four different types of displacements were allowed: (i) translational displacement of a single bead, (ii) pivot rotation, (iii) translation of the entire chain, and (iv) slithering move, in order to accelerate the examination of the configurational space.²⁶ The probability of different trial moves was weighted to enable single-particle moves 20 times more often than the other three. Initially the beads were randomly placed in the box and an equilibrium simulation of typical 5×10^4 trial moves/bead was performed. The proceeding production run comprised 1×10^6 passes. During the production run, in the most concentrated His5 system, the rms displacement of the chain was at least ten times the box length. The distributions of $\langle R_g^2 \rangle^{1/2}$ and $\langle R_{ee}^2 \rangle^{1/2}$ were investigated in order to conclude that the simulations were ergodic.

Electrostatic Properties

His5 is a relatively highly charged protein with linear charge density of $0.11 e/\text{\AA}$ (taking both negatively and positively charged residues into account); hence electrostatic interactions and the charge sequence are thought being of importance, especially in the low ionic strength regime. Since the major group of amino acids in His5 are titrable, the net charge varies between $+15$ to $-5 e$, but His5 is cationic over a wide pH range. At $\text{pH} = 7$ the net charge is $\sim 5 e$. The isoelectric point, determined from the simulated average protonation state of the amino acids, is approximately $10.5 e$ and it is in good agreement with previous reported values.^{27–29} Figure 1 shows the charge distribution of His5 at $\text{pH} = 7$, $\text{IS} = 10 \text{ mM}$ and 150 mM respectively, obtained from MC simulations utilizing the in house software Faunus.²³ The charge distributions are simulated at constant pH and the charge state of the titrable amino acids are allowed to respond to the local chemical environment. Hence, the interactions in the solution give rise to an effective pKa. His5 acquires higher net charge in 150 mM ionic strength in comparison to 10 mM . This since the salt screens the electrostatic interactions and protonation of nearby basic residues is enabled, allowing higher net charge. It is mainly the histidines in the sequence that is the origin of the varying net charge. In addition, due

to charge regulation, the protonation state of the histidines varies depending on its position in the chain.

Results and discussion

Structural properties of His5 have been investigated in aqueous solutions with varying ionic strength and concentrations by utilizing SAXS and Monte Carlo simulations. Both are systematically compared.

Single molecule scattering - The form factor

This section describes the scattering of His5 solutions at physiological IS (140 mM NaCl), 10 mM Tris pH 7 and low peptide concentration (1 mg/mL). Under these conditions, the locations and conformations of the peptides are not correlated, and any interferences between macromolecules are averaged out to zero. The molecular mass from the SAXS data, which provides an indication of monodispersity, was determined from $I(q = 0)$ obtained from the Guinier approximation, restricted to $qR_g < 0.8$, and from the pair distance distribution function ($P(r)$) obtained with GNOM.³⁰ The two values of $I(q = 0)$ were consistent, 0.00279 cm^{-1} and 0.00281 cm^{-1} respectively. From $I(q = 0)$ and the specific volume of His5 ($v_p = 0.7023 \text{ cm}^3/\text{g}$ obtained from Sednterp,³¹) the molar mass of His5 was determined to be approximately 3000 g/mole, in excellent agreement with the value 3036 g/mole calculated from the amino acid sequence. This indicates that His5 indeed is monomeric.

The radius of gyration R_g was determined from both $P(r)$, $R_g = 13.8 \pm 0.04 \text{ \AA}$, and from the Guinier approximation (13.3 \AA). R_g from $P(r)$ is obtained from the entire scattering q range, while R_g from the Guinier plot is restricted to $qR_g < 0.8$ for unfolded proteins. R_g of His5 assessed through $P(r)$ is most likely more reliable since R_g from the Guinier method is known to be less appropriate for extended conformations and tends to underestimate R_g .³² The function $P(r)$ also gives access to the maximal extension of the protein: D_{max}

= 47 Å. Figure 2 shows scattering profiles, Kratky representations, and $P(r)$, calculated from Monte Carlo simulations using the in-house package Molsim³³. The effects of three model parameters; equilibrium bead distance r_0 , Figure 2(a)(b)(c), bead radius $R_{i,j}$, Figure 2(d)(e)(f), and van der Waals attraction reflected in ε , Figure 2(g)(h)(i), between non-connected beads, were studied to get a physico-chemical insights of the SAXS data and His5. A model used by Skepö et al.³⁴⁻³⁷ was employed as reference, $r_0 = 5$ Å, $R_{i,j} = 2$ Å. The SAXS profiles and $P(r)$ inferred by GNOM, are also included in the figures, in black. Figure 2(j)(k)(l) shows the model parameters yielding good agreement between the experimental and simulated data. In addition, completely uncharged His5, and His5 in the absence of all electrostatic screening without van der Waals attraction are shown for comparison. The featureless measured scattering pattern, caused by a vast ensemble of unfolded conformations, is typical for IDPs. $P(r)$ is normalized such that the integral is unity. The unitless Kratky representation qualitatively assess the overall conformational state of His5 and is given by $(qR_g)^2 I(q)/I(0)$ as a function of qR_g . For globular proteins it displays a bell shaped curve with $(qR_g)^2 I(q)$ decreasing as q^{-2} , while rigid rods show a linear increase of $(qR_g)^2 I(q)$ as a function of q . The plot enhances the large q region and reveals the flexibility/rigidity of the peptide. His5 shows a characteristic representation, typical for an IDP, indicated by the lack of a clear maximum in the curve. Instead the profile is approaching an almost plateau and an upward slope at higher scattering angles with a limiting behavior for $(qR_g)^2 I(q)$ of $\sim q^{0.33} - q^{0.3}$ corresponding to a limited behavior for $I(q)$ of $\sim q^{-1.67} - q^{-1.70}$. Note, it is probably rather crude to discuss a 24 aa small protein as His5 (3 kDa) in terms of scaling law from polymer theory, due to its shortness. However, this indicates that His5 is somewhat more extended than a Gaussian chain ($I(q) \sim q^{-2}$) and behaves as a well solvated polymer with excluded volume accounted for $I(q) \sim q^{-1.70}$. In line with previous calculated scattering profile, $P(r)$ of His5 reveals a non-globular and disordered behavior concluded from the asymmetric shape and extended smooth curvature and tail. The peak around 4.1 - 5 Å and the abruptness for low distances come from the

equilibrium distances between the connected hard spheres in the model. Those features could be eliminated by discarding the statistics between the nearest and the next nearest neighbor. $P(r)$ calculated with GNOM, is forcing $P(r)$ to approach zero at $r = 0$. Figure 2(a) shows the scattering profile obtained from SAXS and the scattering curves calculated from Monte Carlo simulation with varying r_0 , 4.1 - 5.0 Å, and $R_{i,j} = 2$ Å. The calculated scattering profiles decay faster than the SAXS profile for low q . The larger the r_0 the steeper profile. The origin to this trend is due to elongation of the chain as r_0 is increased since the slope of the linear fit of $\ln[I(q)]$ vs q^2 in the Guinier region, yields R_g . However, there is no remarkable effect of r_0 on the stiffness of the chain, observed in Figure 2(b). The profiles for large q is almost similar but somewhat more steep in comparison to the experimental one. Thus, in absence of vdW attraction the simulations yield to stiff configurations. This trend is also observed in $P(r)$, see Figure 2(c). The intrachain distances and D_{max} obtained from simulations tend to be too large. The difference in $\langle R_g^2 \rangle^{1/2}$ when changing r_0 from 4.1 to 5.0 is approximately one Å. In Figure 2(d)(e)(f) r_0 is kept constant (5 Å) and $R_{i,j}$ is varied between 1.3 - 2.0 Å. The scattering profiles in Figure 2(d), the experimental and simulated, agree well with the models $R_{i,j}$ of 1.3 and 1.4 Å, while an increase of the bead size beyond 1.4 yields too extended configurations. The stiffness assessed in Figure 2(e) and $P(r)$ in 2(f) also give reasonable good agreement for those two models. It is however not reasonable to believe that so small residue hard sphere radius accurately will represent the excluded volume of an amino acid. In Figure 2(g)(h)(i) r_0 and $R_{i,j}$ are kept constant (5 Å, 2 Å) and an additional attractive vdW potential is added. The potential is varied through ε , 0.3 - 0.9 · 10⁴ kJ Å⁶/mole, giving an attractive potential of 0.3 - 0.9 kT at contact. A vdW of 0.9 kT gives a fairly good agreement with the SAXS profile. However, the curve in the Kratky representation is too flat for large qR_g values, whereas an interaction strength of 0.6 kT on the other hand (blue curve) gives the proper stiffness of the chain. Figure 2(j)(k)(l) shows the results from the model parameters yielding good correspondence to the experimental description of His5, $R_{i,j} = 2$ Å, $r_0 = 4.1$ Å, and vdW = 0.6 kT. Those model parameters will be used further on in

the paper. The chosen $R_{i,j}$ provides a realistic contact separation between charge entities, represents an accurate description of the Coulomb interaction and $\text{vdW} = 0.6 \text{ kT}$ gives an appropriate flexibility. r_0 was adjusted to 4.1 \AA to obtain appropriate chain dimensions. A reduced residual plot as a function of q for the form factor is shown in the supplementary. The average of all conformations obtained from simulating completely uncharged His5, and His5 in the absence of all electrostatic screening without vdW are too extended. The difference between no screening and no charge is small due to the short equilibrium distance, the hard sphere repulsion as well as due to the shortness of the peptide. $\langle R_g^2 \rangle^{1/2}$, $\langle R_{ee}^2 \rangle^{1/2}$, and the reduced χ^2 statistics calculated from the comparison with the intensities, obtained from simulations with the various model parameters are shown in Table 1. The presented values should be compared with the value obtained by SAXS and GNOM, 13.8 \AA .

IDPs as His5 adopt a large number of conformations. Besides MC simulations, Flexible-Meccano (FM)³⁸ has also been used in this study to obtain the pictures of these conformations.

The program FM generates a pool of 10 000 possible polypeptide backbones by randomly selecting specific amino-acid conformations from a library of non-secondary structural elements of high-resolution X-ray crystallographic structures. To the peptide backbone the lateral chains were added by SCWRL4.03.³⁹ Furthermore, the pool of possible conformations was subjected to a generic algorithm (GAJOE, ATSAS package) for a selection of subsets of conformations such that the average scattering of these ensembles fits the experimental data (Figure 3(a), 3(b) grey curve). The R_g distributions from the GAJOE selected ensembles, black, from the entire pool generated by FM (red), and MC (blue) are shown in Figure 3(c). The shift towards larger R_g in the distribution of the selected ensemble indicates that His5 is slightly more extended than predicted from amino-acid-specific backbone dihedral angles obtained from non-secondary structural sequences accessed from high-resolution X-ray crystallographic structures. The average of the curves obtained using FM (see red curve in Figure 3(a), 3(b)) does not fit perfectly the experimental curve. Some rather extended

conformations must be selected by GAJOE in order to obtain a better agreement with the experimental curve (see gray curve in Figure 3(a), 3(b)). On the contrary, the average of the conformations determined by MC is in good correspondence with the experimental curve.

Effect of protein concentration at high ionic strength

Figure 4 depicts the intensities scattered by the His5 solutions in 10 mM Tris buffer pH 7 and 140 mM NaCl, with varying concentrations. There is a slow variation/depression with the concentration at low q , indicative of presence of repulsive interactions, and at 7 mg/mL there is a clear repulsion between the proteins. However, the inset shows that the variation of $I(q = 0)$ with concentrations between 1 to 5 mg/mL is very small and the differences between the extrapolated intensity to zero concentration and the measured intensity at 1 mg/mL are within error bars. In addition, simulations show that there are minor deviations between one single chain and 1 mg/mL. Hence, the error by taking 1 mg/mL as the form factor is within the statistical error of the intensities. Thus, His5 1 mg/mL solution is indeed a good approximation of an ideal solution and gives the form factor.

From MC simulations under physiological conditions, no significant conformational changes of His5 were observed in the studied concentration range (1.1 mg/mL - 7.4 mg/mL), quantified by $\langle R_g^2 \rangle^{1/2}$ and $\langle R_{ee}^2 \rangle^{1/2}$. For the most dilute system $\langle R_g^2 \rangle^{1/2} = 13.78 \text{ \AA}$ std = 2.58 \AA and $\langle R_{ee}^2 \rangle^{1/2} = 33.89 \text{ \AA}$ std = 11.58 \AA , which could be compared to the contour length 106.5 \AA . Hence, the preferable extension of His5 seems to be 30% of the maximum, $\langle R_{ee}^2 \rangle^{1/2} / R_{contour}$. The absence of concentration effects is probably because the free energy associated with changing the conformations is much higher than that associated with changing the relative locations of neighboring chains. No selfassembly of His5 in the studied concentration range was detected from simulations.

Effect of ionic strength

Intramolecular interactions

Ionic strength is an important entity in many systems. Here the electrostatics are treated through a screened Coulomb approach and the addition of salt is taken into account via the screening length κ^{-1} . As the salt content is reduced both intra- and intermolecular electrostatics have to be accounted for. It is difficult to study the effect of ionic strength on the intramolecular distances with SAXS since the scattering profile is affected by the intermolecular interactions at low IS. Even at very low concentration (1 mg/mL), a depression in the profile for low q is present. Figure 5(a) shows the form factor for IS = 10 mM and 150 mM respectively obtained from simulations. The black curve corresponds to the scattering from 1 mg/mL His5, 10 mM Tris, and 140 mM NaCl. The simulated profiles for systems with IS = 10 and 150 mM respectively, are almost indistinguishable. Thus it is reasonable to believe that the intrachain distances are unaffected by IS at pH = 7. In addition, Figure 5(b) showing the partial structure factors (see equation 10), reveals that it is mainly the structure from the neutral residues that is building up the structure of the protein. The upper blue/grey dotted curves represent the structure factor from negatively charged residues, next dotted curves to the positively charged residues, and the full lines to the neutral. The single chain could be represented by an uncharged bead-model if the chain dimensions were tuned. However, the intermolecular interactions would only contain an excluded volume and an entropic part and thus would fail in describing the solution structure at low ionic strength.

The effect of IS on the intrachain dimensions of His5 is further quantified in terms of $\langle R_g^2 \rangle^{1/2}$ and $\langle R_{ee}^2 \rangle^{1/2}$, shown in Figure 6(a), where $\langle R_g^2 \rangle^{1/2}$ is red and $\langle R_{ee}^2 \rangle^{1/2}$ blue vs IS. There are minor effects on $\langle R_g^2 \rangle^{1/2}$ and $\langle R_{ee}^2 \rangle^{1/2}$ of IS and a decrease of $\langle R_g^2 \rangle^{1/2}$ and $\langle R_{ee}^2 \rangle^{1/2}$ with 3.3 % and 1.2 % respectively was observed when altering the salt content, 0 - 300 mM. The effect on P(r) of increased screening is shown in 6(b) and here it is shown that an IS of 10 mM yields more extended configurations in comparison to a more screened system

although the effect is insignificant. The screening length in 1:1 10 mM salt solution is $\sim 30 \text{ \AA}$, which is slightly smaller than the chain dimensions and, hence, the peptide could to some extent be regarded as unscreened. Upon addition of salt to the solution, the screening length decreases and reaches $\sim 8 \text{ \AA}$ and it is solely the residues situated nearby that will interact electrostatically, i.e. short-ranged electrostatic interactions. However, no significant changes were observed even though the screening length was set to values much larger than the chain dimensions. There is always a balance between the electrostatic interactions and the chain entropy when minimizing the free energy expressed in terms of $\langle R_{ee}^2 \rangle^{1/2}$. Assuming an ideal chain, entropy favors a more compact structure due to larger number of possible conformations it can visit in comparison to an extended chain. Thus, even though there is a repulsion between equally charged residues, entropy favors less extended conformations.

Intermolecular interactions

Figure 7(a) shows the scattering intensities for the most concentrated samples, 7 mg/mL His5, solutions with varying NaCl concentrations pH = 7. It is mainly the effect of the intermolecular interaction that is significant in the scattering profile, i.e. depression of the curve in the low q region, reflecting the osmotic compressibility of the solution. Upon decreasing IS the repulsion in the system increases and for IS = 10 mM there is a strong repulsion between the peptides. For this concentration the average interpeptide distance is of the same order of magnitude as the electrostatic interactions and the repulsion increases significantly. Already at IS = 80 mM (green curve), the approximate IS of saliva, the intermolecular electrostatic interactions are effectively screened out and resembles IS = 150 mM (blue curve), as predicted from simulations, here red and grey refer to 50 and 10 mM IS, respectively. In addition, simulated curves for the highest and lowest salt concentrations are shown as dotted lines in Figure 7(b). The peak position of the scattering profile and the structure calculated from simulations of the system with IS = 10 mM, coincide fairly well. Thus, it is reasonable to believe that the interpeptide interactions are accurately captured in

this model. Figure 7(c) shows the partial structure factors calculated from simulations, at high and low IS, and it is clearly shown that the structure factor from neutral amino acids is dominating. The presence of peaks in the structure, in the low salt system, is signifying repulsive interactions between the proteins. Due to the absence of peaks in the high salt system, one can conclude that the origin of the major repulsion is of electrostatic nature. The structure of the neutral residues is determined by the charged residues, to which they are connected within the chain.

Effect of concentration - low ionic strength

Figure 8(a) shows the SAXS scattering profile for varying protein concentrations in a 10 mM Tris buffer, at pH 7. The red curve refers to 2.4 mg/mL, blue 4.7 mg/mL, and green 6.8 mg/mL and the symbols/dotted curves refer to simulations. There is a clear effect of concentration on the profiles and the peak positions are captured well by the simulations. Depending on concentration, the peak positions are shifted in q . The higher the concentration the more shift towards higher q , which indicates smaller average separation between the proteins. From the simulated radial distribution functions between the center of mass of the chains, no inter-protein distance correlations were observed. Figure 8(b) shows distances at which there is a 50 % probability for two center of masses for varying IS vs concentration. At high and intermediate IS there is almost no effect of concentration on the interprotein distances, whereas at low IS there is a clear effect (red curve), due to electrostatic interactions. The dependence of concentration on the intermolecular distances is connected to the second virial coefficient. The higher the concentration the smaller the distances between the center of masses of the proteins. IS = 60 mM and 150 mM, respectively, behaves more as an ideal solution with no concentration effect in this regime. This phenomenon is further demonstrated in Figure 8(c), showing the distances where it is 50 % probability for two center of masses for different concentrations vs IS (inset vs κ^{-1}). There is an exponentially decaying relationship between the distances and salt content and the higher the salt concentration,

the smaller interprotein distances and eventually the distances reaches a plateau at a salt concentration of 150 mM. The effect of concentration also decreases with increasing IS, as discussed above. The remaining concentration effect at high IS comes from the excluded volume of the proteins. In the inset, showing the effect of the screening length, the domination of the electrostatic effect is further manifested. Figure 8(d) shows the running coordination number obtained from simulations, i.e. $\rho \int 4\pi r^2 g(r) dr$, where ρ is the number density. The x -values are normalized with $\langle R_g^2 \rangle^{1/2}$. Red curve refers to 2.4 mg/mL, blue 4.7 mg/mL, and green 6.8 mg/mL protein concentration. The distance within two $\langle R_g^2 \rangle^{1/2}$ is excluded from the solution, hence the intermolecular interaction could be modeled with hard spheres. Beyond two, the probability for two chains to be within the distance increases, and the effect of concentration is observable. There is a 50 % probability for two center of masses within 5.0 , 4.2 and 3.8 r/R_g respectively, going from the lowest to the highest concentration.

Effect of charge distribution

In an aqueous solution with only monovalent ions, a chain with smeared charge distribution, as a polyelectrolyte, expands compared to its unperturbed conformations due to the intramolecular electrostatic repulsion. However, the charge distribution of a peptide is not evenly distributed but rather depends on the amino acid sequence (aa). Thus, the electrostatic intrachain interactions could either give rise to a contraction, to an expansion or no effect at all. In order to determine the effect of the aa sequence and charge distribution on the average conformation and statistical dimensions of His5, the net charge of the chain was evenly distributed and compared with previous model. This is of special interest since His5 is known to bind multivalent ions and thereby altering the charge distribution. However, as the net charge (estimated to 5 e at pH 7 and IS = 150 mM) of the peptide chain is distributed along the chain (0.19 e /residue), the intramolecular distances are virtually not affected by the charge distribution in high IS regime at pH = 7 as long as the net charge is kept constant, $\langle R_g^2 \rangle^{1/2}$ increases 3 % and $\langle R_{ee}^2 \rangle^{1/2}$ 5 %. This since the distribution of

His5 is natively rather homogeneously distributed but also due to its shortness and thereby less statistical possible conformations. In addition, simulations show that the interpeptide interactions are not affected by the charge distribution. Instead it is the net charge of the peptide that is important in the studied concentration regime and the higher the net charge, the more repulsion between the peptides. This indicates that His5 behaves more or less as a small homogeneously charged polyelectrolyte in this regime, which opens up the possibility to investigate the effect of net charge on the conformations of His5 due to bound multivalent ions. His5, which is a metalloprotein, binds various transition metals such as Fe and Zn. This ability may significantly modulate its antifungal function. It is suggested in the literature that His5 binds up to ten equivalents of Fe^{3+} and four of Zn^{2+} ,⁴⁰ leading to a total net charge increase of 30 and 8 e , respectively. In addition, there is a strong correlation between peptide cationicity and antimicrobial activity.⁴¹ The effect of the increased net charge was investigated by calculating the scattering function, $P(r)$ profile, which is sensitive to the symmetry, and quantified in terms of $\langle R_g^2 \rangle^{1/2}$ and $\langle R_{ee}^2 \rangle^{1/2}$. Figure 9 shows the calculated scattering function (a), the Kratky representation (b), and $P(r)$ (c) from simulations and SAXS/GNOM. Black refers to the experimentally obtained scattering (pH = 7, IS = 150 mM), blue to previous model (heterogeneously distributed charges), grey homogeneous charge distribution (0.19 e /residue), green net charge of 13 e (0.5 e /residue), and red refers to net charge of 35 e (1.3 e /residue). From the scattering function, Kratky representation, and $P(r)$ it is obvious that His5 becomes more extended and stiff as the net charge increases, due to repulsion between the residues. $\langle R_g^2 \rangle^{1/2}$ increases with 20 % and 69 % and the extension of the maximum, $\langle R_{ee}^2 \rangle^{1/2}/R_{contour}$, is 41 % and 61 %, upon increasing the net charge with 8 and 30 e . Thus, if the multivalent ions are evenly distributed over the residues of the protein, there is a significant conformational extension to be expected.

Conclusions

The aim of this study was to investigate His5 solution structure as a function of protein concentration and ionic strength, using SAXS, MC, and Flexible-Meccano/GAJOE. At pH 7, at low and high concentration of NaCl, His5 is monomeric and shows scattering profile characteristic for IDPs with extended disorder. It behaves as a well solvated polymer and adopts a large number of conformations. Two approaches have been used to attain the pictures of these conformations. Flexible-Meccano has been utilized to assess the conformations in high IS regime and the results were compared with MC simulations. The conformations obtained by Flexible-Meccano, gave rise to slightly too contracted conformations. From this we conclude that His5 adopts conformations more extended than the average of all accessible conformations. At low IS MC simulations were used to analyze the conformations of His5 and it was concluded that there are no significant conformational changes of His5 between high IS and low IS set by a 1:1 salt, although the protein to some extent could be regarded as unscreened at low IS.

The second aim of this study was to define a coarse-grained model validated by SAXS-results and respond to the question: Is it possible to use a coarse-grained model based on the primitive model to understand the physico-chemical properties of His5 and to capture the effect of electrostatic interactions? Comparison between experimental and simulated results have shown that indeed it is possible to model His5 using a bead-necklace model, based on the primitive model. At high salt concentrations it is mainly the balance between the excluded volume effect and vdW attraction that determines the agreement with the intensity curves. At low salt concentration, the peak positions in the experimental and calculated scattering profiles are located around the same q . This indicates that the intermolecular distances as well as the interactions are well described in the simulations. The intermolecular interactions in the low salt concentration system can be captured by replacing the heterogeneous charge distribution with a smeared charge distribution, or even utilizing a simple hard sphere model, based on the net charge. Regarding the conformational properties of the chain, as the radius

of gyration, there is a small effect using the simpler model. Of course the short-ranged electrostatic interactions within the chain will affect the internal structure of the peptide upon smearing the net charge. Moreover, there is a recognised hypothesis that the anticandidacidal properties of His5 is due to conformational changes of the peptide when binding multivalent ions such as zinc and iron; in this study we have shown that there are no evidence of conformational changes due to electrostatic screening effects. However, a large effect of the protein net charge was observed on the conformational properties of His5. Further studies are directed into conformational changes with respect to binding of divalent ions explicitly as well as the interaction with lipid bilayers.

Acknowledgments

We acknowledge financial support from: Organizing Molecular Matter (OMM), Vinnova, the Vinnmer program, The Royal Physiographic Society in Lund, Per-Eric and Ulla Schybergs Foundation, and the Crafoord foundation. The simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at the center for scientific and technical computing at Lund University (LUNARC). We are grateful to Dr. Petra Pernot at ESRF, Grenoble for providing assistance in using beamline BM29.

References

- (1) Higham, S. M.; Cate, J. M.; Ship, J. A.; Smith, P. M.; Tenovuo, J.; Whelton, H. In *Saliva and Oral Health*, 3rd ed.; Edgar, M., Dawes, C., O'Mullane, D., Eds.; BDJ Books, 2004.
- (2) Dunker, A. K. et al. *J Mol Graph Model* **2001**, *19*, 26–59.
- (3) Wright, P. E.; Dyson, H. J. *J Mol Biol* **1999**, *293*, 321–331.

- (4) Liu, J.; Faeder, J. R.; Camacho, C. J. *Proceedings of the National Academy of Sciences* **2009**, *106*, 19819–19823.
- (5) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *Journal of molecular biology* **2004**, *337*, 635–645 0022–2836.
- (6) Puri, S.; Edgerton, M. *Eukaryotic cell* **2014**, *13*, 958–964 .
- (7) Ruissen, A.; Groenink, J.; Helmerhorst, E.; Walgreen-Weterings, E.; van 't Hof, W.; Veerman, E.; van Nieuw Amerongen, A. **2001**, *356*, 361 – 368 .
- (8) Bennick, A. *Critical Reviews in Oral Biology & Medicine* **2002**, *13*, 184–196.
- (9) Wróblewski, K.; Muhandiram, R.; Chakrabartty, A.; Bennick, A. *European Journal of Biochemistry* **2001**, *268*, 4384–4397.
- (10) Siqueira, W.; Margolis, H.; Helmerhorst, E.; Mendes, F.; Oppenheim, F. *Journal of Dental Research* **2010**, *89*, 626–630.
- (11) Brewer, D.; Hunter, H.; Lajoie, G. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire* **1998**, *76*, 247–256.
- (12) Raj, P. A.; Marcus, E.; Sukumaran, D. K. *Biopolymers* **1998**, *45*, 51–67.
- (13) Henriques, J.; Cragnell, C.; Skepö, M. *Journal of Chemical Theory and Computation* **2015**, *11*, 3420–3431, PMID: 26575776.
- (14) McQuarrie, D. A. *Statistical Mechanics*, 1st ed.; University Science Books, Sausalito, California, 2000.
- (15) Henriques, J.; Cragnell, C.; Skepö, M. *Journal of Chemical Theory and Computation* **2015- resubmitted (April)**,
- (16) Iovino, M.; Falconi, M.; Marcellini, A.; Desideri, A. *The Journal of Peptide Research* **2001**, *58*, 45–55.

- (17) Kurut, A.; Henriques, J.; Forsman, J.; Skepö, M.; Lund, M. *Proteins: Structure, Function, and Bioinformatics* **2014**, *82*, 657–667.
- (18) Raj, P. A.; Edgerton, M.; Levine, M. J. *Journal of Biological Chemistry* **1990**, *265*, 3898–3905.
- (19) Raj, P. A.; Soni, S. D.; Levine, M. J. *J Biol Chem* **1994**, *269*, 9610–9619.
- (20) Skepö, M. *The Journal of Chemical Physics* **2008**, *129*, 185101.
- (21) Skepö, M.; Linse, P.; Arnebrant, T. *Coarse-grained modeling of proline rich protein 1 (PRP-1) in bulk solution and adsorbed to a negatively charged surface*; Journal of Physical Chemistry B American Chemical Society, 2006; Vol. 110:24, s. 12141-12148.
- (22) Evers, C. H.; Andersson, T.; Lund, M.; Skepö, M. *Langmuir* **2012**, *28*, 11843–11849 .
- (23) Lund, M.; Persson, B.; Trulsson, M. *Biol. Med* **2008**, *3:1*.
- (24) Nozaki, Y.; Tanford, C. *Enzyme Structure*; Methods in Enzymology; Academic Press, 1967; Vol. 11; pp 715 – 734.
- (25) Allen, M.; Tildesley, D. *Computer Simulations of Liquids*; Oxford University Press: New York, 1987.
- (26) *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*; Oxford University Press: New York, 1995.
- (27) Helmerhorst, E. J.; van't Hof, W.; Breeuwer, P.; Veerman, E. C.; Abee, T.; Troxler, R. F.; Amerongen, A. V.; Oppenheim, F. G. *J. Biol. Chem.* **2001**, *276*, 5643–5649.
- (28) Nikawa, H.; Fukushima, H.; Makihira, S.; Hamada, T.; Samaranayake, L. P. *Oral Diseases* **2004**, *10*, 221–228.
- (29) Kurut, A.; Henriques, J.; Forsman, J.; Skepö, M.; Lund, M. *Proteins: Structure, Function, and Bioinformatics* **2014**, *82*, 657–667.

- (30) Svergun, D. I. *Journal of Applied Crystallography* **1992**, *25*, 495–503.
- (31) <http://www.jphilo.mailway.com/default.htm>.
- (32) Bernadó, P. **2010**, *39*, 769–780.
- (33) P., L. MOLSIM, Lund University, Sweden 2013.
- (34) Skepo, M. *Journal of Chemical Physics* **2008**, *129*.
- (35) Skepo, M.; Lindh, L.; Arnebrant, T. *Zeitschrift für physikalische Chemie* **2007**, *221*, 21–46.
- (36) Skepo, M.; Lindh, L.; Arnebrant, T. *Zeitschrift für physikalische Chemie* **2007**, *221*, 21–46.
- (37) Skepo, M.; Linse, P.; Arnebrant, T. *Journal of Physical Chemistry B* **2006**, *110*, 12141–12148.
- (38) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J.-r.; Jensen, M. R.; Segard, S.; Bernadó, P.; Charavay, C.; Blackledge, M. *Bioinformatics* **2012**, *28*, 1463–1470.
- (39) Georgii G. Krivov, R. L. D., Maxim V. Shapovalov <http://dunbrack.fccc.edu/scwr14>,
- (40) Puri, S.; Li, R.; Ruszaj, D.; Tati, S.; Edgerton, M. *Journal of Dental Research* **2015**, *94*, 201–208.
- (41) Mochon, A. B.; Liu, H. *PLoS Pathog* **2008**, *4*, e1000190.

Figure Legends

Figure 1. Charge distribution of His5 in 1:1 salt concentration 10 mM (red) and 150 mM (blue) respectively pH = 7, calculated using MC simulations. The net charge is about 5 e and varies slightly depending on IS.

Figure 2. Measured form factor (black), simulated intensities, Kratky representations, and the $P(r)$, with varying model parameters, IS = 150 mM, pH = 7. (a)(b)(c) shows the conformational effect of equilibrium distance, r_0 , (d)(e)(f) effect of bead size, $R_{i,j}$, (g)(h)(i) effect of van der Waals strength between non connected beads, varying ϵ . (j)(k)(l) The model used in the paper.

Figure 3. (a) Comparison of the His5 SAXS form factor (pH = 7 10 mM Tris, 140 mM NaCl)(black) with the average of the SAXS curves associated with all conformations selected by GAJOE (gray), and coarse - grained Monte Carlo simulations (MC) (blue). (b) Kratky representation of His5 obtained with SAXS (black), MC (blue), and Flexible Meccano (red), and GAJOE (grey). (c) Distribution of R_g values obtained from MC simulations (blue), conformations of ensembles (200 conformations per ensemble) selected by the GAJOE algorithm to fit the experimental form factor of His5 (black), and the distributions of R_g from the entire pool of conformations generated by FM (red).

Figure 4. SAXS intensities for 140 mM NaCl, 10 mM Tris, at pH 7. Red refers to 1 mg/mL, blue to 2.4 mg/mL, green to 4.7 mg/mL, and black to 7.4 mg/mL Histatin 5.

Figure 5. (a) Form factor determined from SAXS, 1 mg/mL His5, IS =150 mM. The simulated curves were obtained from a single chain, IS = 10 mM and 150 mM, respectively. (b) The partial structure factors obtained from the positive, negative and neutral beads respectively, building up the chain in the model.

Figure 6. (a) Simulated $\langle R_g^2 \rangle^{1/2}$ (red) and $\langle R_{ee}^2 \rangle^{1/2}$ (blue) as an effect of IS. The largest estimated uncertainties are $\sigma(\langle R_g^2 \rangle^{1/2}) = 0.1 \text{ \AA}$ and $\sigma(\langle R_{ee}^2 \rangle^{1/2}) = 0.4 \text{ \AA}$. (b) The pair distance distribution functions obtained from simulations with varying IS and SAXS/GNOM (black) IS = 150 mM.

Figure 7. (a) Scattering intensities obtained by SAXS, varying IS, pH = 7, and 7 mg/mL His5. Grey IS = 10 mM, red IS = 50 mM, green IS = 80 mM, and blue IS = 150 mM. (b) Blue curve refers to 7 mg/mL His5 and IS = 150 mM and grey to 7 mg/mL His5 and IS = 10 mM respectively. The lines with shapes show the corresponding simulated systems. (c) The partial structure factors that corresponds to the positive, negative and neutral beads, obtained from simulations with low and high salt content.

Figure 8. (a) SAXS curves from different His5 concentrations at low IS (10 mM Tris, pH 7), red 2.4 mg/mL, blue 4.7 mg/mL, and green 6.8 mg/mL. The symbols/dotted lines refer to simulated profiles from similar systems. (b) The distances obtained from radial distribution functions, $g(r)$, at which there is a 50 % probability for two center of masses vs concentration. Red curve refers to IS = 10 mM, black IS = 60 mM, and blue IS = 150 mM. (c) The distances obtained from $g(r)$ at which there is a 50 % probability for two center of masses vs IS. The inset shows the 50 % probability distances vs screening length. (d) Running coordination number obtained from MC simulations of His5 solutions, IS = 10 mM, red 2.4 mg/mL, blue 4.7 mg/mL, and green 6.8 mg/mL.

Figure 9. Evaluation of the effect of charge distribution/net charge of His5 on the intramolecular distances of the protein. Black = obtained from SAXS via GNOM, blue = heterogeneous charge distribution, grey = 0.19 e/residue, green = 0.5 e/residue, and red = 1.3 e/residue (a) Scattering function (b) The Kratky representation (c) $P(r)$.

Table

Table 1: $\langle R_g^2 \rangle^{1/2}$ and $\langle R_{ee}^2 \rangle^{1/2}$ ^a obtained from simulations with varying model parameters and the sums of the squared residuals shown as the reduced χ^2 statistic. Equilibrium distance between the beads, $r_0 = 4.1 - 5.0$ Å, bead radius, $R_{i,j} = 1.3 - 2.0$ Å, and vdW = 0 - 0.9 kT.

Model parameters						
r_0 (Å)	$R_{i,j}$ (Å)	vdW (kT)/comment	$\langle R_g^2 \rangle^{1/2}$ (Å)	$\langle R_{ee}^2 \rangle^{1/2}$ (Å)	χ_{red}^2	
4.1	2.0	0	15.09	38.13	13.5	
4.4	2.0	0	15.44	38.86	15.1	
4.6	2.0	0	15.66	38.89	15.2	
4.8	2.0	0	16.04	40.20	28.4	
5.0	2.0	0	16.13	40.21	73.8	
5.0	1.3	0	13.66	32.99	0.9	
5.0	1.4	0	13.99	33.82	0.5	
5.0	1.8	0	15.49	38.22	15.1	
5.0	1.9	0	15.93	39.69	27.1	
5.0	2.0	0	16.13	40.21	73.8	
5.0	2.0	0.9	13.80	33.09	2.9	
5.0	2.0	0.6	14.94	36.80	13.8	
5.0	2.0	0.3	15.66	38.93	27.0	
5.0	2.0	0	16.13	40.21	73.8	
4.1	2.0	0.6	13.78	33.89	0.2	
4.1	2.0	no charge	14.67	37.94	16.4	
4.1	2.0	no screening	16.21	42.15	34.4	

^aLargest estimated uncertainties are $\sigma(\langle R_g^2 \rangle^{1/2}) = 0.09$ Å and $\sigma(\langle R_{ee}^2 \rangle^{1/2}) = 0.38$ Å.

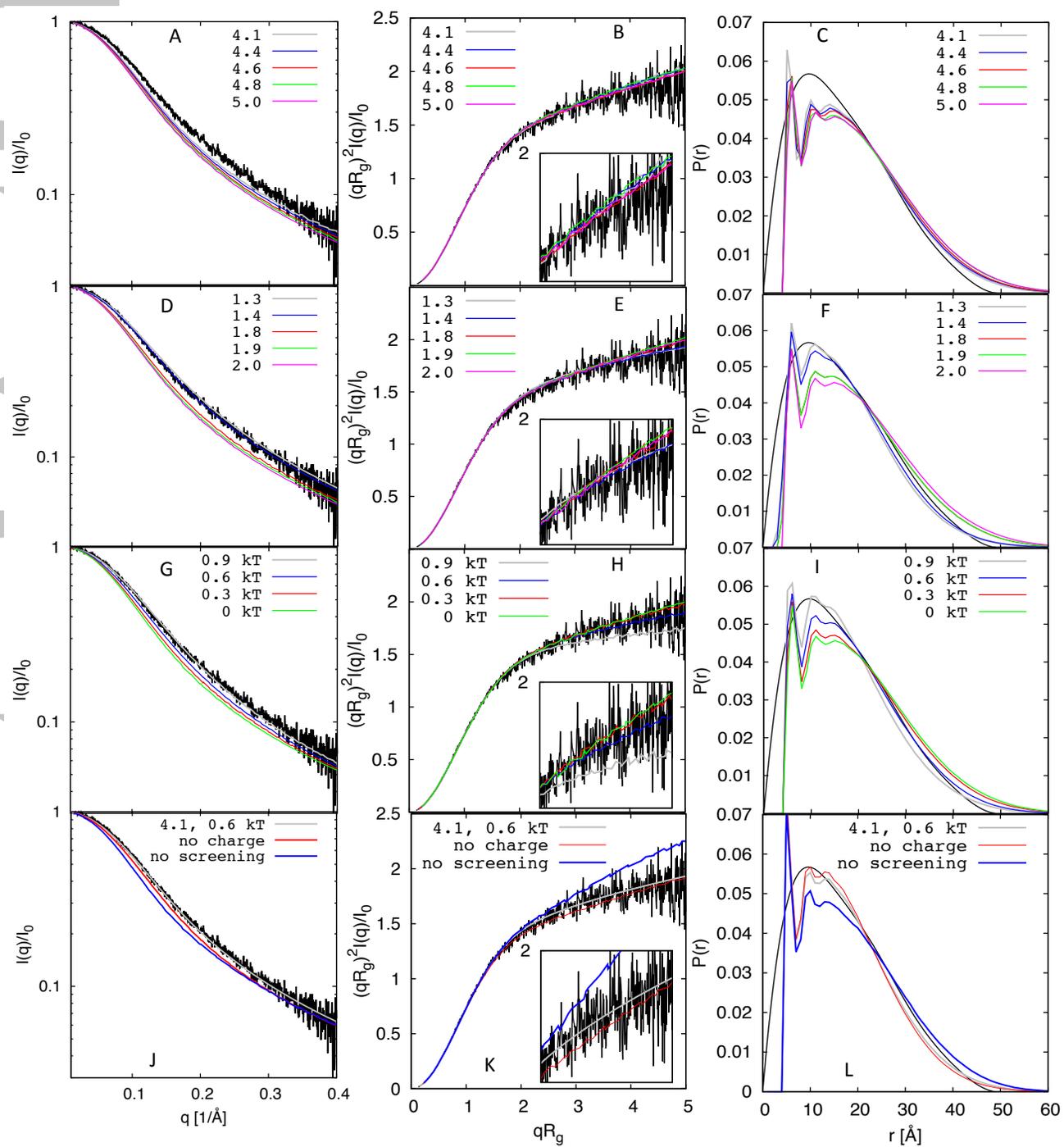


Figure 2

Accepted Article

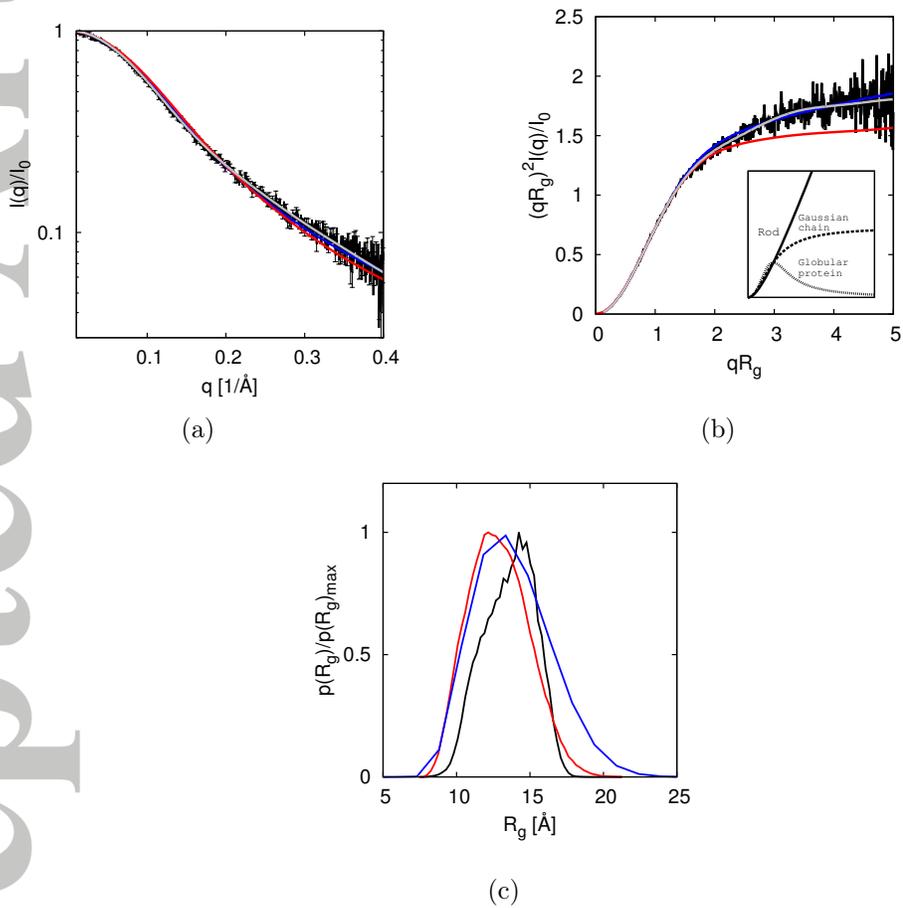


Figure 3

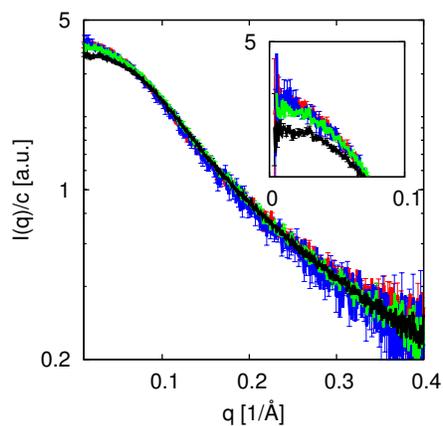
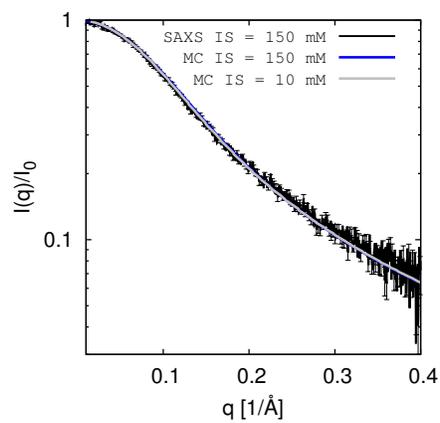
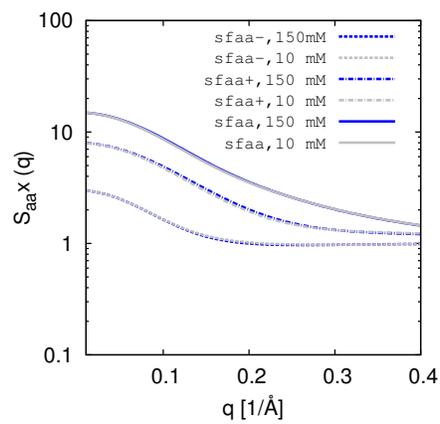


Figure 4



(a)



(b)

Figure 5

Accepted Article

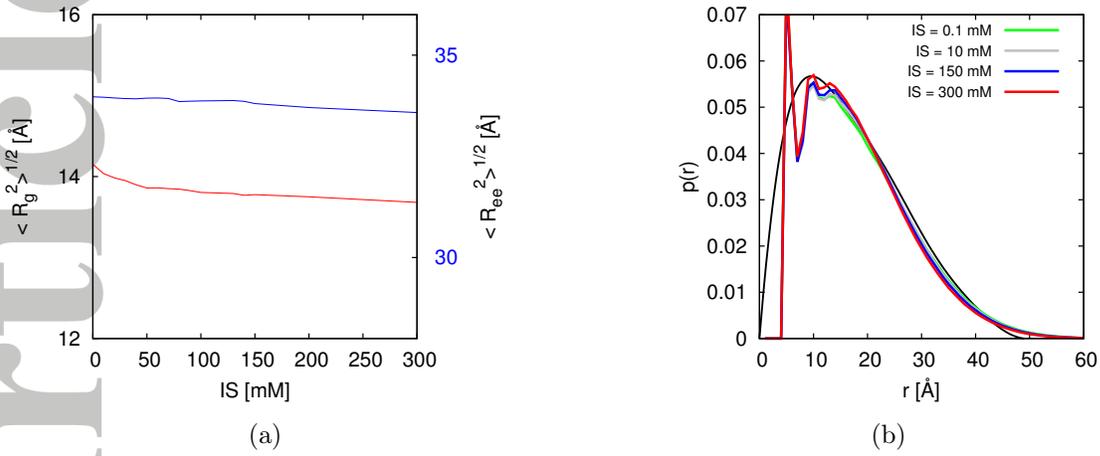


Figure 6

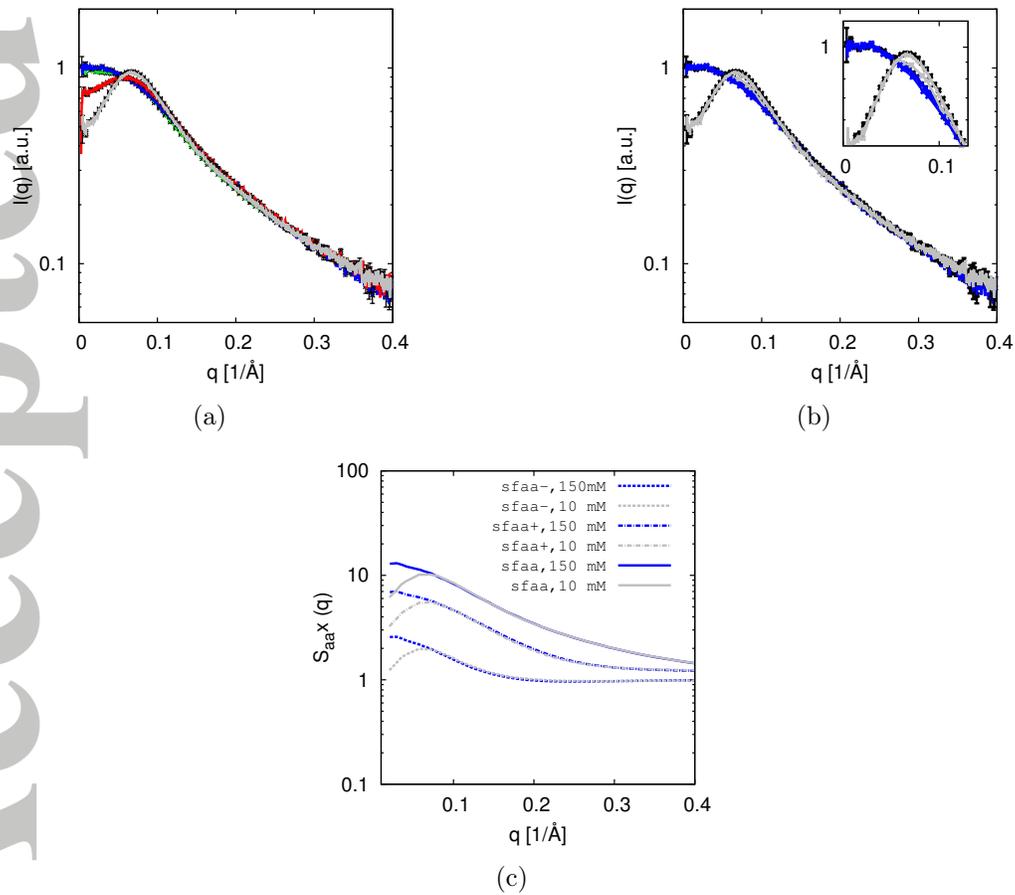
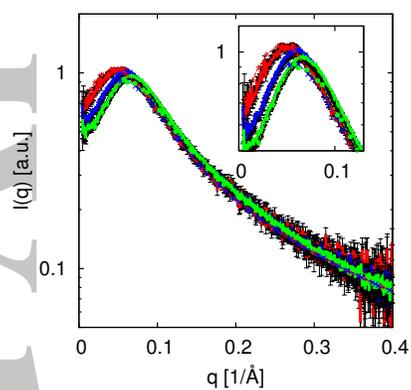
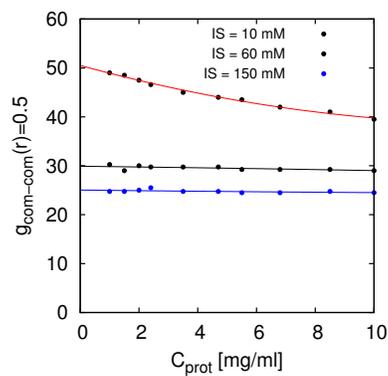


Figure 7

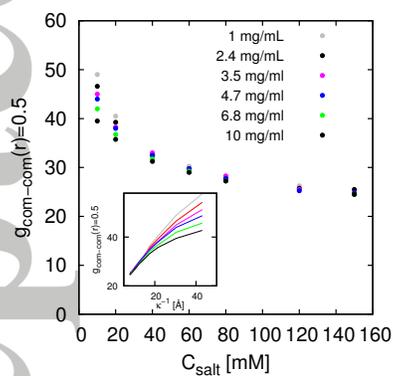
Accepted Article



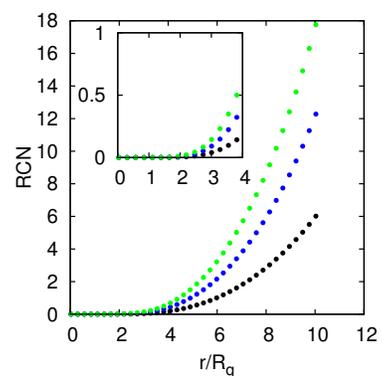
(a)



(b)



(c)



(d)

Figure 8

Accepted Article

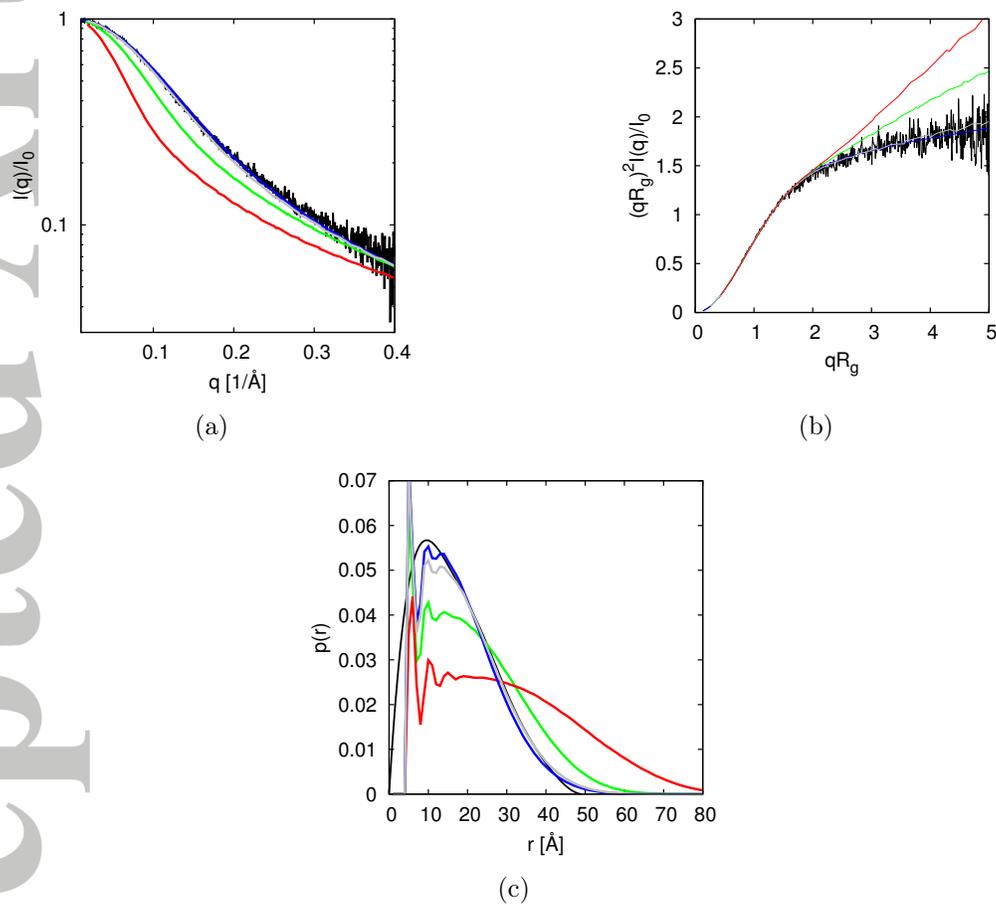


Figure 9