Journal of Chromatography A, 1217 (2010) 1343–1353

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

3D structure-based protein retention prediction for ion-exchange chromatography

Florian Dismer*, Juergen Hubbuch

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Science, University of Karlsruhe (TH), 76131 Karlsruhe, Germany

ARTICLE INFO

Article history:
Received 23 September 2009
Received in revised form 3 December 2009
Accepted 22 December 2009
Available online 4 January 2010

Keywords:
Protein dynamics
Adsorption energies
Binding orientation
Lysozyme
Retention volume prediction
Ribonuclease A

ABSTRACT

The interest in understanding fundamental mechanisms underlying chromatography drastically increased over the past decades resulting in a whole variety of mostly semi-empirical models describing protein retention. Experimental data about the molecular adsorption mechanisms of lysozyme on different chromatographic ion-exchange materials were used to develop a mechanistical model for the adsorption of lysozyme onto a SP Sepharose FF surface based on molecular dynamic simulations (temperature controlled NVT simulations) with the Amber software package using a force-field based approach with a continuum solvent model. The ligand spacing of the adsorbent surface was varied between 10 and 20 Å. With a 10 Å spacing it was possible to predict the elution order of lysozyme at different pH and to confirm *in silico* the pH-dependent orientation of lysozyme towards the surface that was reported earlier. The energies of adsorption at different pH values were correlated with isocratic and linear gradient elution experiments and this correlation was used to predict the retention volume of ribonuclease A in the same experimental setup only based on its 3D structure properties. The study presents a strong indication for the validity of the assumption, that the ligand density of the surface is one of the key parameters with regard to the selectivity of the adsorbent, suggesting that a high ligand density leads to a specific interaction with certain binding sites on the protein surface, while at low ligand densities the net charge of the protein is more important than the actual charge distribution.

1. Introduction

Chromatography is the most important technique in downstream processing of biotechnological products. Although is has been used extensively over the past decades, the understanding of the underlying fundamental mechanisms is still limited, especially when it comes to adsorption mechanisms on a molecular level. Studies about the effects of mobile phase composition on the retention behaviour of proteins – especially in ion-exchange chromatography – resulted in semi-empirical models aiming at the prediction of retention time under changing experimental conditions. One of the most popular models in this regard is the steric mass-action model (SMA) by Brooks and Cramer [1], followed by other models, such as the available area model by Bosma and Wesselingh [2,3] or the stoichiometric displacement model by Rounds and Regnier [4–6]. All these models require the determination of protein and adsorbent specific parameters. Although these parameters are all empirically defined, they have an underlying physical relevance. For example, the SMA model uses the three parameters ionic capacity, describing the number of available ligands in the adsorbent resin, the steric hindrance factor, describing the amount of space the bound protein blocks on the adsorbent surface, and the characteristic charge, describing the number of interaction sites between the protein and the adsorbent surface.

The predictive power of the SMA model – as well as most other approaches – is based however on the underlying assumption that the binding mechanism remains unchanged [7–11]. With a change in the mobile phase pH, the charge distribution on the surface of the protein changes, which in return might change protein surface interactions, specifically the site on the protein mainly responsible for binding [12]. As models fail to account for changes on the protein molecular level, there are approaches such as quantitative structure–property relationship (QSPR) to include these effects [13–15]. These approaches use structural descriptors together with a statistical evaluation of experiments to establish a link between protein properties and the retention behaviour. Yang et al. [13] for example treated proteins at different pH values as distinct molecules in order to allow retention time prediction at different pH. Nevertheless, the training of this model required a data set of >250 experiments.

Besides these semi-empirical approaches, there are also mechanistical models available to describe protein retention on ion-exchange materials. Roth and Lenhoff [16] investigated the adsorption of lysozyme onto a charged surface *in silico* by calculat-

* Corresponding author. Tel.: +49 721 608 6236; fax: +49 721 608 6240.
E-mail address: florian.dismer@kit.edu (F. Dismer).

0021-9673/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2009.12.061

ing the electrostatic and van der Waals energies for the interaction already in 1993. Due to the limitations in the computational speed at that time, they used a sphere representation of lysozyme with the net charge located at the center of the sphere for most of their calculations. In their studies, this showed to be a valid assumption, allowing them to demonstrate the effects of ionic strength on the affinity of lysozyme. In 1991, Stahlberg et al. [17] published a paper dealing with the relationship between the Gibbs free energy and the retention factor for proteins on charged surfaces. In this work, the interaction between the protein molecule and the adsorbent surface was simplified by using two charged surfaces differing in their charge density. In 1992, he extended his model by addition of van der Waals forces [18]. In 1999, he wrote a detailed review article [19] about retention models in ion chromatography, mainly focusing on small molecules, but also including a chapter about charged macromolecules. With the increase of computational speed during the last years it is now possible to perform simulations of the dynamics of proteins in solution rather than static calculations. This development clearly has a potential to reach a higher level of understanding and insights into protein–surface interactions [20–24]. In this paper modelling approaches were built on a detailed mechanistical understanding of the adsorption behaviour of lysozyme onto a SP Sepharose FF adsorbent surface at varying pH determined experimentally and published earlier [12,25]. These experimental results were used to construct an adequate adsorbent surface model of SP Sepharose FF *in silico* to perform molecular dynamic simulations and to access adsorption energies at different pH and for different proteins.

In the presented approach, the interaction between an adsorber surface (SP Sepharose FF) and two proteins (lysozyme and ribonuclease A) was characterized by MD simulations. The simulations were temperature controlled (NVT type simulations) and without boundary conditions, meaning that the surface had a limited size and was not periodically repeated. The force-field used for the simulations was the *ff03* force-field developed by Duan et al. [26], which is a modified version of the *ff99* force-field by Wang et al. [27], both general force-fields for MD simulations with proteins and nucleic acids. All simulations were performed using a generalized Born continuum solvent model initially implemented by Onufriev et al. [28]. In a first step, a simplified model for the adsorber surface was constructed, meaning that the polymer backbone (namely agarose) was not included in the model. The charge carrying ligands of the adsorbent were designed according to their chemical structure. To fix the ligands in space, which is necessary for building a surface-like structure, a positional restraint energy in the form of:

$$E_i^{restraint} = k(\Delta x_i)^2 \quad (1)$$

was applied to i atoms, where k is the weight of positional restraint energy (which was set to +1.0 kcal/(mol Å)) and Δx is the difference between the actual Cartesian coordinates of the restrained atom i and its reference position. For each ligand molecule, the atom that would normally be covalently attached to the polymer backbone of the adsorbent surface and the next two atoms were restrained, and the above described restrained energy was applied to each of these atoms throughout the whole simulation. Although the surface of a real adsorbent particle would probably not be planar from a macroscopic point of view, on the scale of a single protein molecule this assumption should be acceptable. This surface model allows for a certain flexibility of the ligands due to their chemical structure, accounting for experimental findings by DePhillips et al. [29] who found a significant influence of the spacer length on the retention behaviour of proteins. Three different ligand spacings were chosen: 10, 15 and 20 Å(defining the distance to the next ligand in x and y direction in the surface plane). After generating the surface, it was a highly ordered structure. For the starting structure of the surface to be more realistic and to validate whether or not the starting conformation of the surface had any influence on the course of the simulation, a short MD simulation was run with the surface to generate three different starting structures for the actual MD simulation with surface and protein.

In a second step, these three surfaces were used to generate three surface–protein ensembles. The center of mass of the protein molecule (either lysozyme or ribonuclease A) was brought above the center of the surface, and the distance between the protein and the nearest ligand was adjusted to 5 Å. A MD simulation was performed (20 ps) for each ensemble in order to extract the different energies of the ensemble: bond-, angle-, dihedral-, van der Waals-, electrostatic- and restraint-energy. To be able to calculate the energies caused by adsorption for different protein orientations, the MD simulations were repeated for altogether 62 different, systematically varied protein orientations to screen the whole protein surface for possible interaction sites (for detailed information see the materials and methods section). For such a screening it was of great importance, that (a) the protein does not change its orientation in the course of the simulation, and (b) the protein does not change its distance to the surface. Therefore the backbone atoms (imido-N, C_α and carbonyl-C) of the protein were restrained in the same way as described above for the surface ligands. The MD simulation results for all three different starting structures for each protein orientation were compared to determine the effects of the starting conformation on the MD simulation. The different energies were plotted versus the protein orientation (defined by two rotation angles) into interaction plots which revealed favourable and unfavourable binding sites and were correlated with experimental data about binding orientations. Average electrostatic interaction energies were calculated from the gathered data and correlated with retention experiments. All MD simulations were performed for different pH values. To estimate the effect of the pH on the charge distribution and net charge of the protein, protonation states of all charged amino acids were calculated and considered in the MD simulations. An empirical correlation between electrostatic interaction energy and retention behaviour was then used to estimate the retention behaviour of a second model protein (ribonuclease A) only based on the 3D structure of the molecule.

2. Materials and methods

2.1. Isocratic and gradient elution experiments

Hen egg white lysozyme (L-6876) and ribonuclease A from Sigma (St. Louis, MO, USA) was dissolved in the respective working buffer at various concentrations. All salts including 1 M NaOH for pH adjustment were purchased from Merck (Darmstadt, Germany). The adsorbent material analyzed (SP Sepharose Fast Flow) was obtained from GE Healthcare (Uppsala, Sweden). To determine the pH-dependent strength of interaction of lysozyme with SP Sepharose FF, retention experiments were conducted. Lysozyme solutions were prepared at pH 5–12 using 10 mM citrate buffer (pH 5), phosphate buffer (pH 7), carbonate buffer (pH 9), glycine (pH 11) or NaOH (pH 12). The columns used were 1 ml prepacked columns (Atoll, Germany), separation was performed on an Akta system purchased from GE Healthcare (Uppsala, Sweden). Sample loading (240 μl, 3 mg/ml lysozyme) was followed by two column volumes wash with buffer A (10 mM, buffer type depending on pH). The elution was done with a linear gradient from 0% to 100% buffer B (10 mM buffer + 0.5 M NaCl) over 30 column volumes. The conductivity at the point of elution was used to determine the strength of interaction between adsorbent and protein.

2.2. Adsorber surface construction

The chemical structure of the ligand is available on the homepage of the manufacturer (http//www.gehealthcare.com):

Agarose–$OCH_2CHOHCH_2OCH_2CH_2CH_2CH_2SO_3^-$

In a real adsorbent particle, the ligand is coupled to the agarose matrix via an oxygen atom. Due to the fact, that the polymer matrix was not included in the model, an additional C-atom was added to represent the coupling point. The 3D structure of the ligand was created using ANTECHAMBER, a tool provided with the AMBER 9 package, a commonly used package for molecular mechanics [30,31]. The ligands where then distributed evenly to give a quadratic surface of at least 100 Å size in both directions (the actual size slightly changed with different ligand spacings), which is approximately 2-times the size of lysozyme. Three different spacings between ligands were used to generate three different surfaces: 10, 15 and 20 Å (the spacing defines the distance to the next ligand in *x*- and *y*-direction within the surface plane). To keep the ligands in a layer and to avoid a breakdown of the surface during the simulation, the Cartesian coordinates of the first three atoms of each ligand (C_3, O and C_1) were restrained by a harmonic potential of the form mentioned above (see Eq. (1) with k = +1.0 kcal/(mol Å)) during all simulations. This means that only the part of the ligand, that would be attached to the polymer backbone of the adsorbent is fixed in space, but the part carrying the charge (which is the part interacting with the protein) remains flexible, resulting in a time-dependent charge distribution of the surface.

2.3. Protein structure preparation

For the dynamic simulations with lysozyme (PDB-ID: 132L) [32,33] and ribonuclease A (PDB-ID: 1FS3) [33,34] at different pH values, the internal pK_a of lysine, arginine, histidine, glutamic acid and aspartic acid where calculated using the PCE (protein continuum electrostatics) web tool [35]. The charge of each amino acid was then determined individually using the Henderson–Hasselbalch equation:

$$\mathrm{pH} = \mathrm{p}K_a + \log \frac{c(A^-)}{c(HA)} \quad (2)$$

and assigned to the residues using the LEAP tool (also part of the AMBER 9 package). Before the protein structure was used, a short energy minimization (1500 iteration steps) was performed with a continuum solvent model (generalized Born solvation model, Amber parameter: igb = 5, for details read [28]) to avoid general problems with the energetics of the protein structure, such as close contacts between atoms, abnormal torsion angles, etc.

2.4. Ensemble construction and simulation design

To get the final ensemble of adsorbent surface and protein molecule, the protein structure was added onto the surface, with the center of mass of the protein being above the center of the surface. The distance between the surface and the protein was adjusted to 5 Å between the protein and the closest SO_3^- group (which in most cases belonged to one of the ligands in the center of the surface). This was done to keep the distance of the protein to the surface equal in all simulations independent of the orientation of the protein. For each protein, 62 different orientations in respect to the adsorbent surface were sampled. This was achieved by stepwise rotating the protein by 30° around the *y*- (*x*- and *y*-axis lie in the surface plane) and the *z*-axis (perpendicular to the surface plane) to screen the whole surface of the protein for possible interaction sites. After each rotation step, the protein was centered and adjusted to the correct distance. To avoid a change of the orientation during the simulation and to avoid a protein movement away from the surface, the coordinates of the protein backbone (imido-N, C_α and carbonyl-C) were restrained with the same energy as the ligand base atoms (see above), and thus the protein was kept at its position, but still having flexible side chains and flexible adsorbent ligands. Prior to the actual simulation run, the energy of the whole ensemble was minimized for another 1500 steps by using a continuum model (generalized Born solvation model, Amber parameter: igb = 5, for details read [28]) to avoid close contacts between the side chains and the ligands which might lead to atypically high energies in the simulation.

2.5. Simulation parameters and computational equipment

The calculations were all done on 64 cores of a HP XC6000 parallel computing cluster with Intel Itanium2 processors with the SANDER tool (also part of the AMBER 9 package).

The force-field used was the *ff03* force-field developed by Duan et el. [26], which is a modified version of the *ff99* force-field by Wang et al. [27]. All performed MD simulations were of the NVT type, meaning that the number of atoms, the volume and the temperature were kept constant in each simulation run. A Langevin temperature control was used [36,37] which uses imaginary atom collisions to control the velocities of the atoms in the system. The collision frequency was set to 1 ps^{-1}. Bond interactions including hydrogen atoms were omitted to reduce the computational cost. All simulations were performed using a generalized Born continuum solvent model initially implemented by Onufriev et al. [28]. The salt concentration of the continuum was set to 20 mM which influences the dielectric constant of the continuum. If not stated differently, the simulation time was 20 ps with time steps of 0.002 ps. Periodic boundaries were not used, which means that the surface had a fixed size and was not periodically repeated. The cut-off distance was set to 60 Å to include long-ranged electrostatic interactions, and no correction term was added for interactions beyond the cut-off range. The following energies were calculated in each MD run: bond-, angle-, dihedral-, van der Waals-, electrostatic- and restraint-energy. As mentioned above, part of the ligand atoms and the protein backbone were restrained with a harmonic potential of +1 kcal/(mol Å).

2.6. Snapshot sampling and averaging

All simulations were performed in triplicates. For MD simulations, the course of the simulation may depend on the starting conformation of the system (see Section 3 and thus for the evolution of the different energies throughout the simulation. To evaluate the effect of the starting structure on the energies calculated during the simulation, three different starting structures for each ensemble were generated. A dynamic simulation of only the surface was done first with a total time of 100 ps starting with a highly ordered surface model (see Fig. 1).

Three different snapshots of the surface were taken between 20, 60 and 100 ps simulation time to generate three different sets of ligand coordinates (with a maximum period of simulation time between the snapshots to allow for a bigger difference between the structures) for different conformational states of the surface. These different surfaces were used to generate the starting structure of three different ensembles per protein orientation, and the energies of all three simulations were averaged later. In some cases, the restraint energy (the energy actually needed to keep the ligands and the protein at their position) for the ensemble simulation for one snapshot was significantly higher for all 62 orientations than for the other two snapshots. As this is a strong indicator for deeper problems with the energy of the system (e.g. close contacts between protein and ligands to the ensemble construction procedure). Only the other two corresponding snapshots were used to

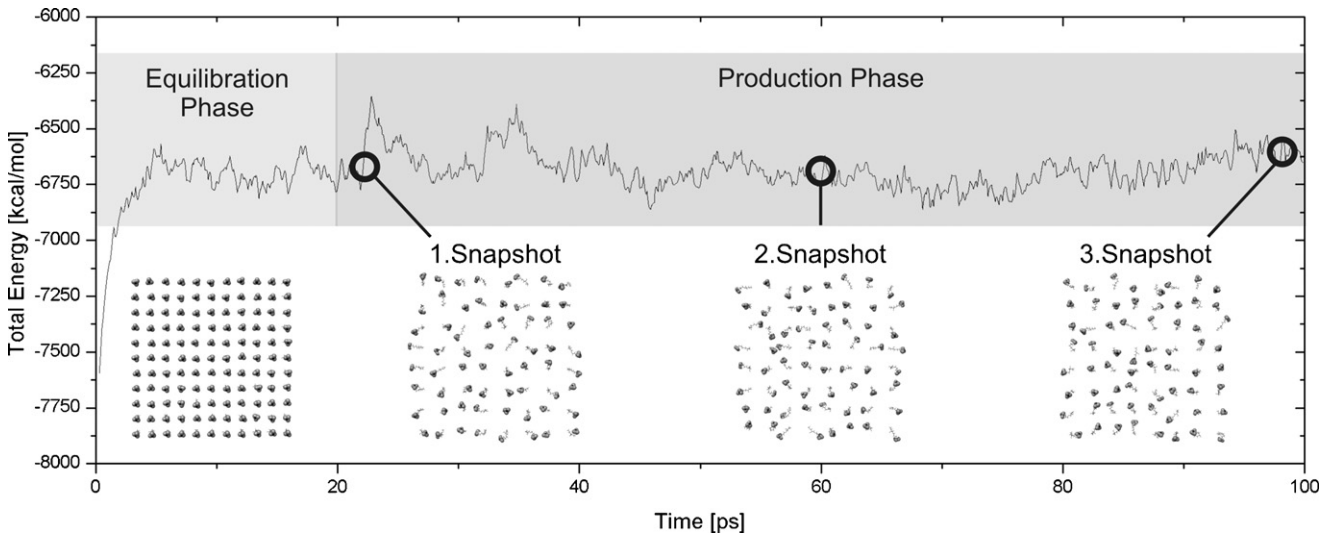


Fig. 1. Plot of the total energy over simulation time for the initial simulation used to generate snapshots for the ensemble generation for the surface with 10 Å ligand spacing.

calculate the average energies. To calculate the interaction energies, reference runs with only the surface and only the protein at a distinct pH were performed in triplicates and averaged. These reference energies were then subtracted from the energies of the ensemble simulation according to:

$$E_{Adsorption} = E_{Ensemble} - E_{Protein} - E_{Surface} \quad (3)$$

The first 10 ps of each simulation were considered to be an equilibration phase, and were thus not used in the calculation of the average energies.

For the correlation of electrostatic interaction energies with retention behaviour, a few assumptions were made:

1. Each protein orientation showing a negative electrostatic interaction energy (which means that there is a net attraction between protein and adsorbent) contributes to protein binding and thus needs to be considered when calculating the average interaction energy.
2. Each protein orientation showing a positive electrostatic interaction energy and thus a net repulsion between protein and surface does not contribute to the average interaction energy (occurred only at pH 12, 10 Å ligand spacing).
3. The more negative the interaction energy of an orientation is, the stronger is the contribution the average interaction energy. The general idea behind this is, that a protein with one strong binding site, which is superior to all other binding sites shows a stronger binding than a protein with many weak binding sites, even if the non-weighted average interaction energy is the same. Thus the interaction energies for the individual orientations need to be weighted before averaging.

The MATLAB 4 grid data method was used to interpolate energies with a stepsize of 5° which were then plotted according to Fig. 2.

The weighting factor for a given orientation can be calculated according to the Boltzmann distribution, which describes the probability of a certain orientation depending on its energy:

$$p_{y,z} = \frac{1}{Z} \times e^{-\beta \times E_{y,z}^{elec}} \quad (4)$$

where $p_{y,z}$ is the probability for a given orientation defined by the two rotation angles y and z, Z is a normalizing factor, $E_{y,z}^{elec}$ is the electrostatic interaction energy of this orientation and β is a constant:

$$\beta = \frac{1}{k_B \times T} \quad (5)$$

where k_B is the Boltzmann constant and T is the temperature. Z can be calculated according to:

$$Z = \sum e^{-\beta \times E_{y,z}} \quad (6)$$

Then an average electrostatic interaction energy $\bar{E}_{y,z}^{elec}$ would be the sum of the weighted electrostatic interaction energies over all

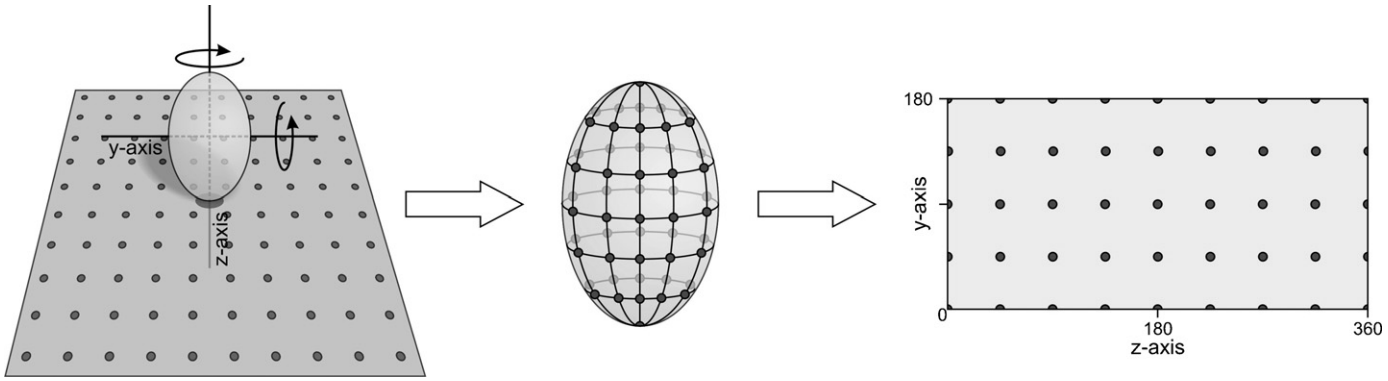


Fig. 2. Scheme of the data flow. By rotating the molecule consecutively around the y- and z-axis the complete protein surface is sampled. Each orientation is specified by a y- and z-angle. These coordinates are then used to plot the energies gathered in the simulations.

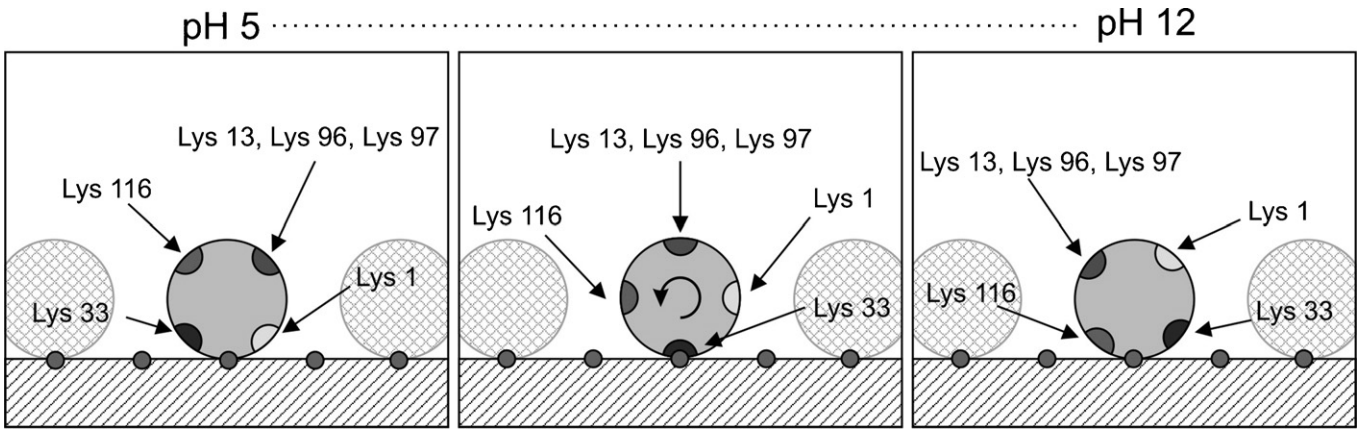


Fig. 3. Changes in the binding orientation of lysozyme on SP Sepharose FF based on experimental data taken from Dismer et al. [12]. At low pH the main binding site was located between lysine 1 and lysine 33. With increasing pH, the molecule turned towards a second binding site between lysine 33 and lysine 116.

orientations that show $E_{y,z}^{elec} < 0$:

$$\tilde{E}^{elec} = \sum p_{y,z} \times E_{y,z}^{elec} \quad (7)$$

3. Results and discussion

3.1. Experimentally determined binding orientations

In a previous publication dealing with binding orientations of lysozyme on different adsorbent surfaces under varying experimental conditions [12] a binding mechanism for lysozyme on SP Sepharose FF was proposed. This binding mechanism is illustrated in Fig. 3.

For low pH the main binding site for lysozyme was located between lysine 1 and lysine 33, including altogether 4 positively and 1 negatively charged amino acid. With increasing pH the lysozyme molecule turned towards a second binding site located between lysine 33 and lysine 116 also consisting of 4 positively and 1 negatively charged amino acid. The driving force for this reorientation was the change of charge distribution on the surface of the protein. All amino acids located in the binding site have slightly different intrinsic pK_a values and are thus loosing their charge at different pH. Also the N-Terminus is located in the first binding site and is the first amino acid that is deprotonated (already below pH 7). For a detailed explanation please refer to [12]. In the present paper these experimental findings were correlated to the dynamic simulation results and subsequently used to determine the most realistic ligand spacing for the investigated systems.

3.2. Surface design and snapshot generation

Fig. 4A shows a model of the surface with a spacing of 10 Å. This structure was used for the generation of three snapshots for the ensemble construction and for reference simulations. The ligands in the simulation are actually not bound to a surface, they are rather just fixed in space and kept at their position by applying a restraint energy. Fig. 4B shows RMS values for the ligand atoms which represent the average distance of an atom relative to its mean position throughout a complete simulation, thus the higher the RMS value, the more movement an atom made during the simulation. The first three atoms (C_3, O and C_1) are the atoms that were restrained at their starting Cartesian coordinates by the harmonic potential, representing a covalent attachment to the surface. These three atoms had the lowest RMS value of all (~0.5 Å). With increasing distance to these atoms, the RMS value increased. There was an obvious connection between the ligand density and the flexibility of the

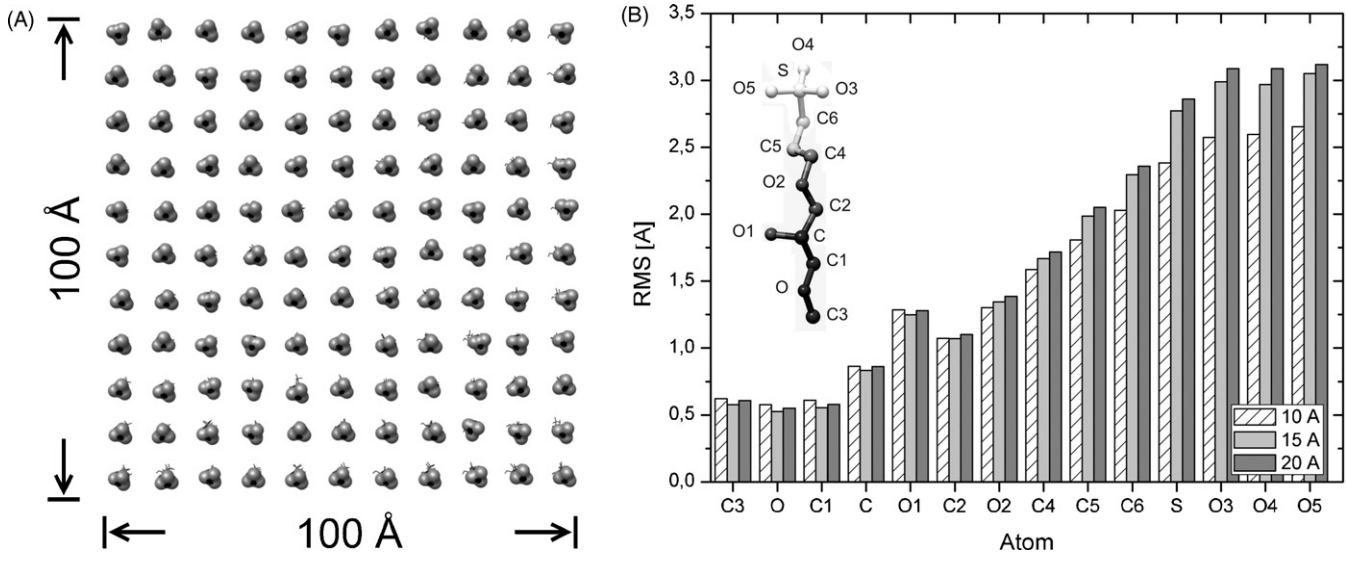


Fig. 4. (A) Model of a SP Sepharose FF surface with a ligand spacing of 10 Å and a total size of 100 Å.SO_3^- groups are colored in gray. (B) RMS values for ligand atoms for three different ligand densities showing their flexibility during a simulation run (green = high flexibility, red = low flexibility).

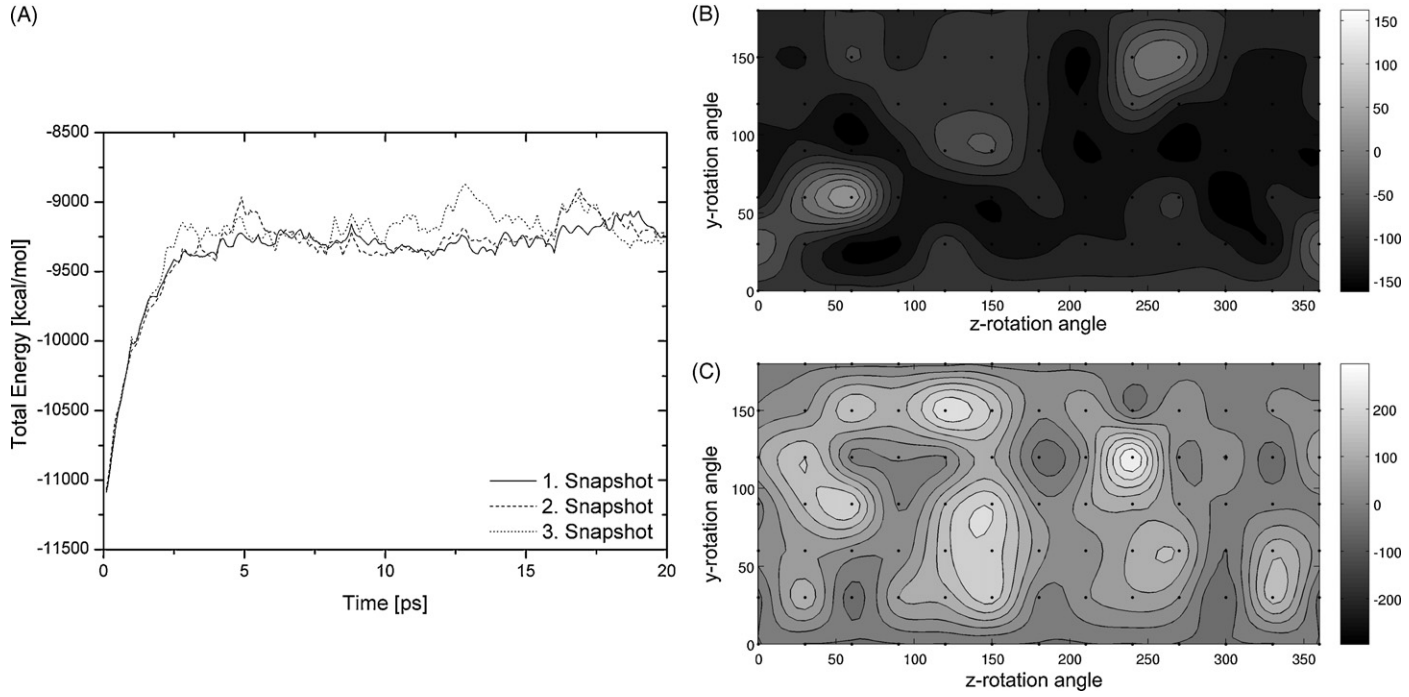


Fig. 5. (A) Total energy for three independent simulations with three different starting structures (10 Å spacing, pH 7, $y = 0^\circ$, $z = 0^\circ$). (B and C) Restraint energy profiles for the first and the third Snapshot of three individual simulations (15 Å spacing, pH 5). The restraint energy profile of the third snapshot reveals problems with the energetics of the third simulation.

ligands. For the 10 Å ligand spacing, the ligands slightly stabilized each other, resulting in lower RMS values. This stabilization was less prominent for increasing ligand spacing. The main reason for restraining three atoms in space rather than one was to avoid a 180° rotation of the ligand facing into the 'surface'.

Fig. 1 shows the proceeding for the generation of three different sets of coordinates for the finale protein–surface ensemble. The first 10 ps of the simulation are considered to be an equilibration phase, in which kinetic energy is added to the system to bring the simulation to a temperature of 300 K. After the equilibration phase the data generation phase follows. In this phase three different sets of coordinates of the surface were being extracted. The pictures below the energy curve in Fig. 1 show that the three snapshots actually represent the same surface only with different conformations and that the surface remained intact throughout the whole simulation.

3.3. Snapshot averaging

Fig. 5A shows the total energy for three different simulation runs performed with three different snapshots (starting conformations) of the same system (10 Å ligand spacing, pH 7, $y = 0^\circ$, $z = 0^\circ$). The data production phase begins after 10 ps. It was obvious that all three simulations fluctuated around a similar average total energy: −9257, −9222 and −9134 kcal/mol, although the course of the simulation was different due to the different starting structures. The averaging of energies during the production phase of each simulation showed a standard deviation of less than 3.5% for all snapshots and all pHs. The materials and methods section describes, that the restraint energy of a simulation can be used as a measure for the quality of the simulation. Fig. 5B shows a typical restraint energy profile for 62 different orientations (on a surface with a ligand spacing of 15 Å at pH 5). The profile was calculated by using reference simulations with the protein and the surface only as described earlier (see Eq. (3)). For most of the simulations, the restraint energy of the 2-component system was lower than for individual 1-component systems. Fig. 5C shows the restraint energy profile for the third snapshot under the same conditions, revealing significant problems with the energies of the system such as close contacts between two or more atoms (the second snapshot is not shown because it was very similar to the first snapshot). Simulation runs with such a high restraint energy profile were not considered in the averaging procedure. This averaging procedure had a standard deviation of below 2% for all pHs. These high restraint energies for some simulations seemed to correlate with the finding, that one or more ligands flipped to the other side of the surface, pointing away from the protein, rather than pointing towards the protein. In these cases the restraint energies would be relatively high as two of the three restrained atoms of these ligands were far away from the position they were supposed to be kept at, and the larger the distance to that position, the higher the restraint energy is. Since these surface configurations did not represent a configuration of a real adsorber surface (the polymer backbone of the adsorber would make such a flipping impossible), we did not include these simulations in any further calculations.

3.4. Binding orientations and ligand spacing effects

The simulation results for the surface with 10 Å ligand spacing and pH values between 5 and 12 are shown in Fig. 6A–E.

The first noticeable finding at pH 5 was, that the electrostatic energy strongly depended on the orientation of the protein: the difference between the lowest energy (−7588 kcal/mol) and the highest energy (−5321 kcal/mol) was roughly 2300 kcal/mol. The second finding was, that there were apparently two sets of orientations, that were unfavourable for a binding event: one set of orientations around $y = 75^\circ$, $z = 100^\circ$ and one set around $y = 115^\circ$, $z = 220^\circ$.

Although these orientations are unfavourable, the electrostatic energy for the interaction was still negative and these orientations had to be considered in the calculation of an average interaction energy. Binding orientations 1 and 2 were the ones with the lowest energy. Note that the orientation 1 was almost exactly the orienta-

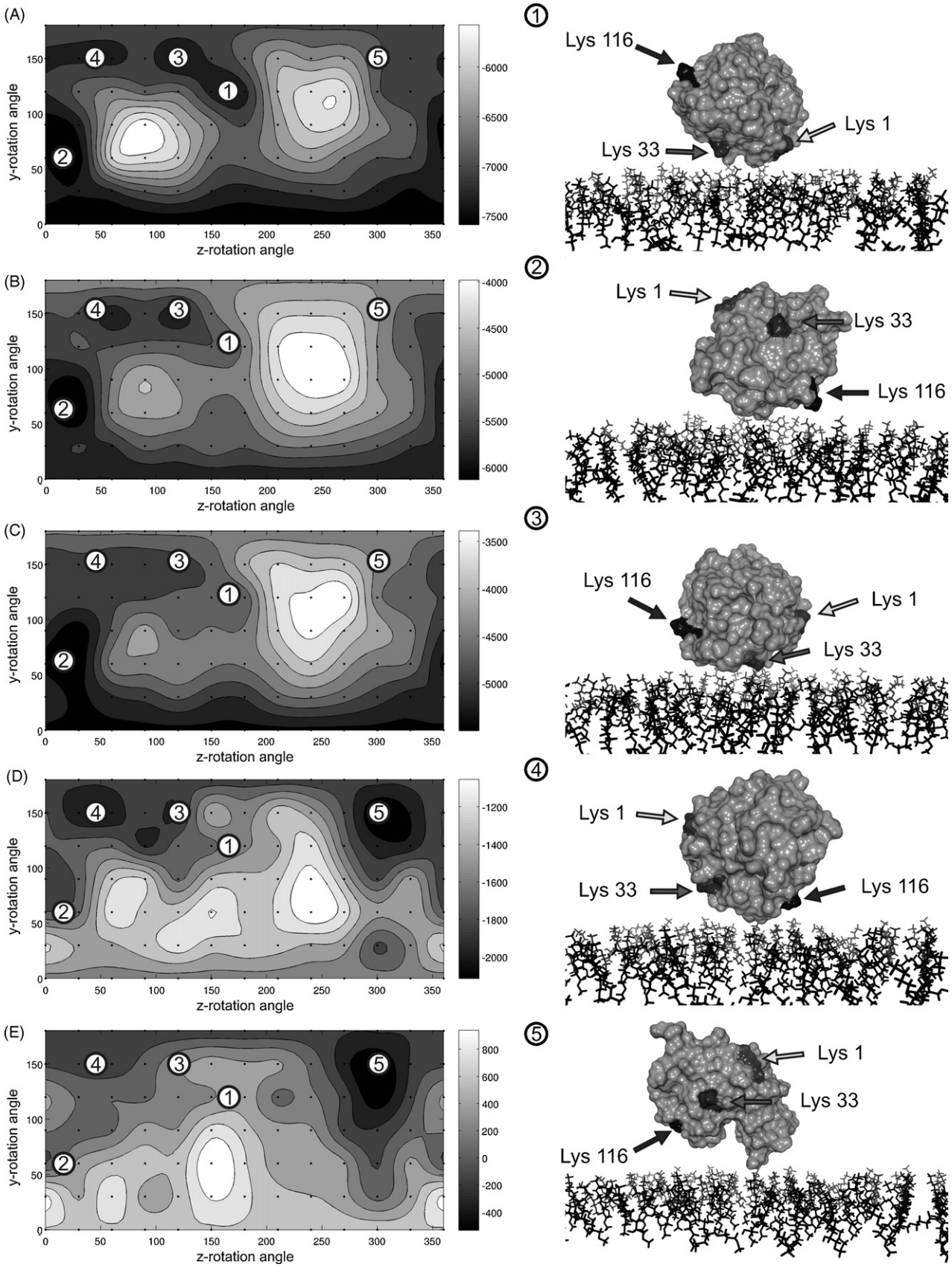


Fig. 6. Electrostatic energy profiles for different pH: (A) pH 5, (B) pH 7, (C) pH 9, (D) pH 11 and (E) pH 12 and a selection of the most favourable binding orientations (1–5).

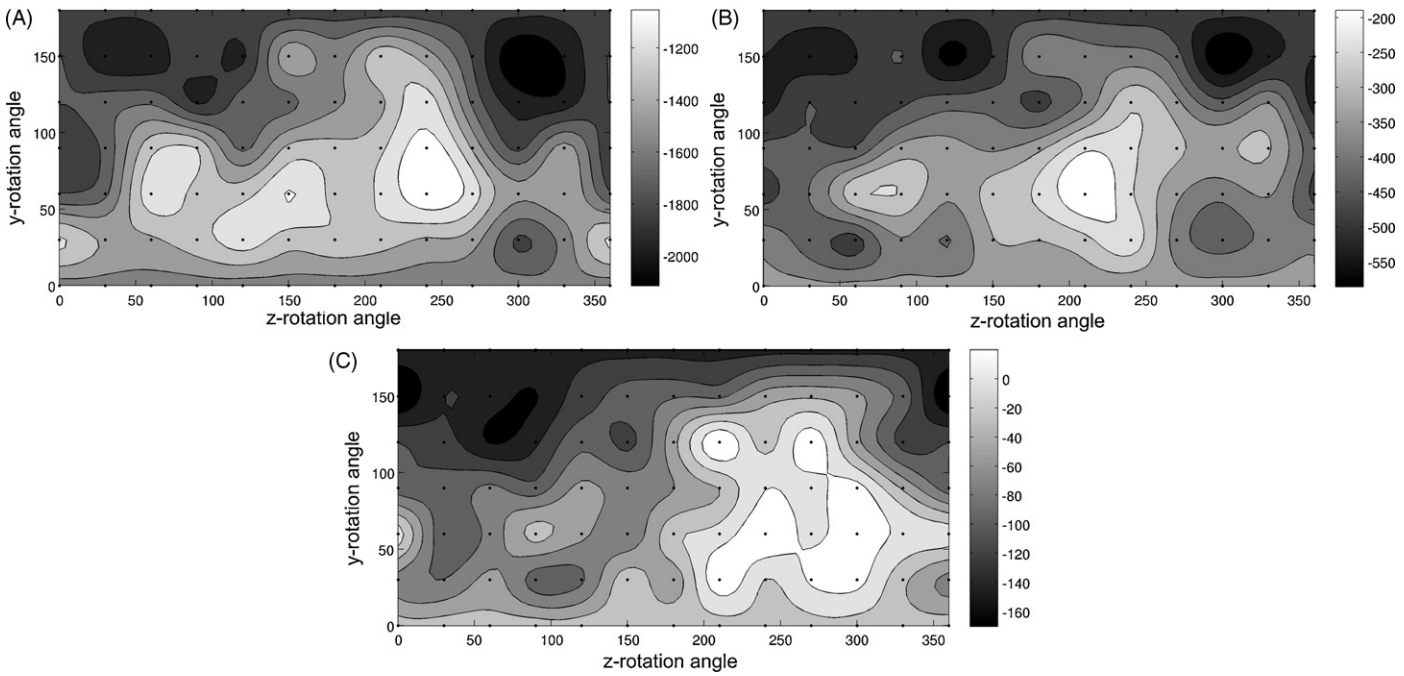


Fig. 7. Effects of (A) 10 Å, (B) 15 Å and (C) 20 Å ligand spacing on the binding orientation of lysozyme here shown at pH 11.

tion that was proposed from the experimental data [12] (see Fig. 5). With increasing pH to pH 7, the total electrostatic energy increased by approximately +1400 kcal/mol, but the difference between the lowest and the highest energy remained roughly 2400 kcal/mol. The first unfavourable binding site became less prominent due to a loss of negative charges at asparagines 52 and 101 located close to the surface at this orientation. At pH 11, new unfavourable binding sites occurred due to changes in the charge distribution. At pH 12 areas of electrostatic repulsion appeared as expected. Nevertheless, there were still regions in the plot showing a negative electrostatic energy, correlating well with the experimental finding that lysozyme showed significant retention beyond its p*I* at pH 11.3 [12].

Note again, that the most prominent binding orientations (no. 4 + 5) were close to the ones that were proposed from experimental data [12] (see Fig. 3).

Fig. 7A–C shows the effects of ligand spacing on both, the binding orientation and the electrostatic energy for the interaction between lysozyme and the surface at pH 11. With increasing ligand spacing, the average electrostatic energy increased from about −1547 kcal/mol to about −97 kcal/mol. Apparently the electrostatic binding becomes also less selective to the charge distribution as the difference between the lowest and the highest energy is reduced from about ∼ 1000 to ∼150 kcal/mol. This finding corresponds and explains reports claiming that ligand density influences not only the maximum binding capacity but also the selectivity of adsorbent materials. Wu and Walters [38] have shown, that with changing ligand density on a silica-based material, the elution order of lysozyme and cytochrome changes.

The key message of this section is, that a whole variety of potential binding sites exists, some with a higher probability than others. Clearly, the binding orientation greatly influences the energy for the adsorption governing the probability of orientation. The simulation results showed that the binding is far more complex than the experimentally determined data suggested, but a good correlation of both experimental and *in silico* data could be found.

3.5. Ribonuclease A as a second model protein

To evaluate the predictive power of the MD simulations, ribonuclease A was used as a second model protein. Ribonuclease is a protein with a similar size (124 residues) and an estimated p*I* of 9.5 resulting in a positive net charge at pH 7. For the simulations, the surface with a 10 Å ligand spacing of the same size as the surface used for the simulations with lysozyme was chosen, representing the system with the best correlation (as can be seen later). 62 different orientations were analyzed and the results were interpolated as described in the materials and methods section. The results are

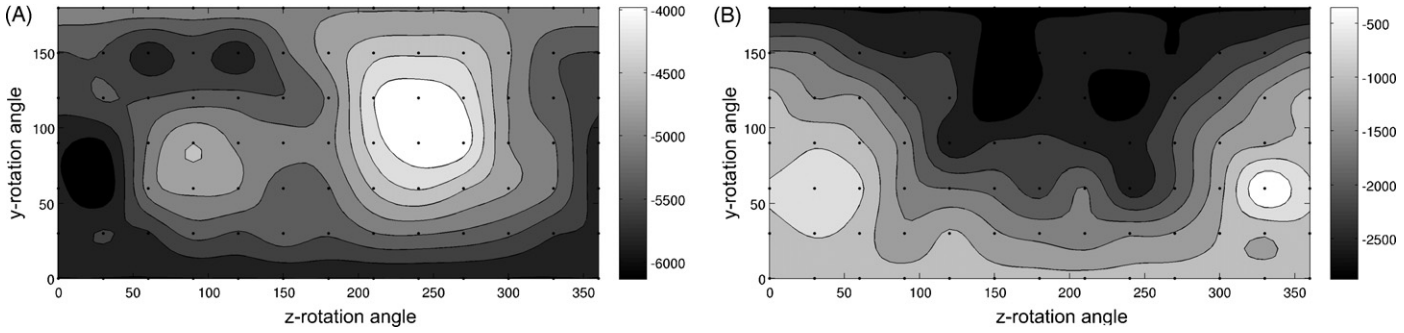


Fig. 8. Simulation results for (A) lysozyme and (B) ribonuclease A on a surface with 10 Å ligand spacing at pH 7.

shown in Fig. 8 in comparison to simulations for lysozyme at pH 7.

The calculated average electrostatic energy at pH 7 was −1950 kcal/mol compared to −5164 kcal/mol for lysozyme, so just by looking at this energy, a significantly lower retention volume was expected for both, isocratic and gradient elution.

3.6. Correlation with elution studies: isocratic elution

In the previous section it was discussed that there are numerous possible binding sites differing in their electrostatic energies. The question arose whether these electrostatic energies correlate not only with the adsorption behaviour but can also be used to predict desorption behaviour. In order to predict the retention of a protein two prerequisites would be needed: (1) a general approach to calculate an energy of adsorption from the simulation results and (2) a correlation between this interaction energy and the retention of a protein. In order to calculate an average electrostatic energy of interaction the assumptions were made, that each orientation with a negative energy can possibly bind and lead to retention, while orientations showing a positive energy would not contribute, and that the lower the energy is, the higher the contribution to the retention (for details see Section 2. To account for both, weighted average energies were calculated as described in the materials and methods section. As also discussed earlier, in the course of this study simulations with three surfaces differing in their ligand spacing were performed. This section will focus on data generated with the 10 Å surface, as they showed the best results. For measuring the affinity of lysozyme at different pH isocratic elution experiments were performed at pH 5, 7, 9 and 11. The results are shown in Fig. 9.

The log(k) versus log(conductivity) plots showed a linear correlation as described elsewhere [39,40]. The two parameters from the isocratic elution experiments (k and κ) were then correlated with a third parameter, the average electrostatic interaction energy obtained from MD simulations by multilinear regression. The generated three-dimensional correlation surface was of the form:

$$\log(k) = a \times \log(\kappa) + b \times \bar{E}^{elec} + c \quad (8)$$

where κ is the conductivity and $\bar{E}^{elec}_{pH}$ is the average electrostatic interaction energy from MD simulations. This correlation could then be used to calculate retention factors k for any given combination of buffer conductivity and interaction energy. Parameters a, b and c were determined with lysozyme data (isocratic elution experiments and MD simulations at different pH) and used to predict k for ribonuclease A at pH 7. The results are shown in Fig. 10A.

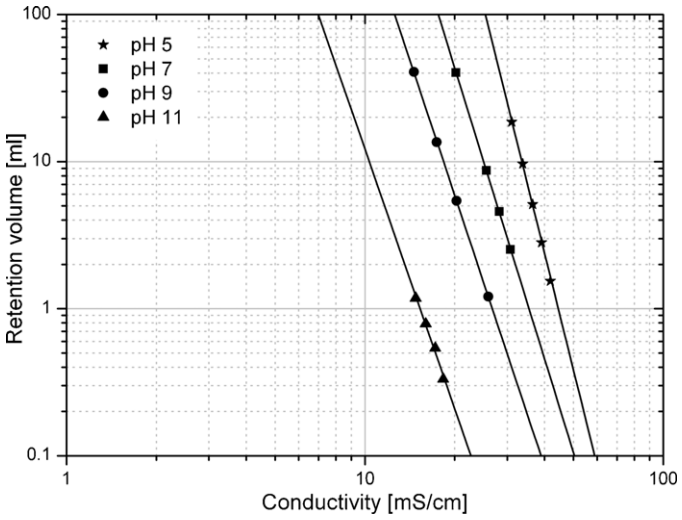
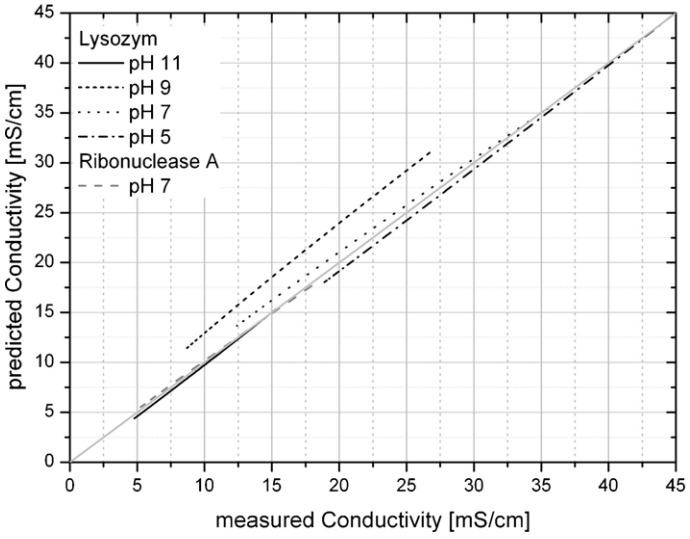


Fig. 9. Isocratic elution experiments with lysozyme at pH 5, 7, 9 and 11.

Measured and predicted conductivities agreed well for all pHs, except for pH 9. The parameters were: $a = -6.867$, $b = -6.932 \times 10^{-4}$ mol/kcal and $c = 6.884$. It should be noted at this point, that these parameters are probably only valid for the exact same MD simulation setup. One possible explanation for that is an inaccurate internal pK_a calculation. There are several ways to estimate internal pK_a values of amino acids within a protein:

- Application of empirical rules.
- Solution of the Poisson–Boltzmann equation to account for influences by other charged amino acids within the structure.
- Calculation of the energy needed to remove a proton from the protonated form of an amino acid within a protein structure.

In this paper, the second approach was used to prepare the protein structure models for the MD simulations. It will be part of future studies to investigate the effect of other approaches on the MD simulation results. Fig. 10B shows the prediction result for ribonuclease A at pH 7 and compares the result to a predic-



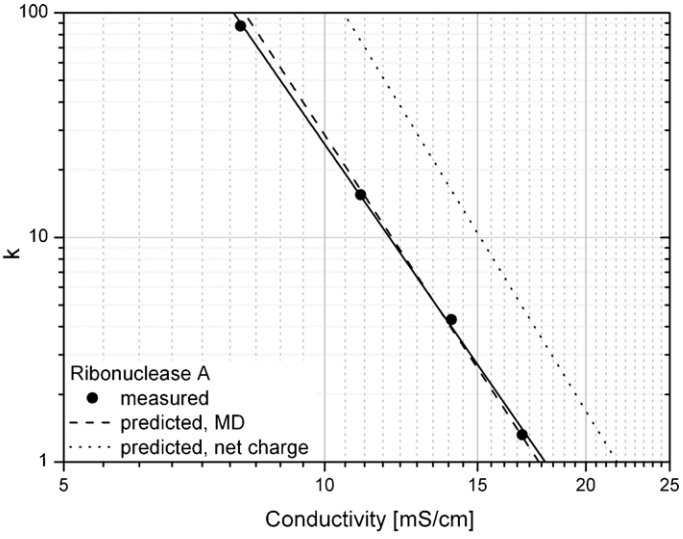


Fig. 10. (A) Predictive power of the multilinear regression. The correlation was good for all pH except pH 9. Ribonuclease A behaviour was also well predicted. (B) Prediction of retention behaviour of ribonuclease A at pH 7 using MD simulations (dashed line) and the net charge (dotted line) of the protein.

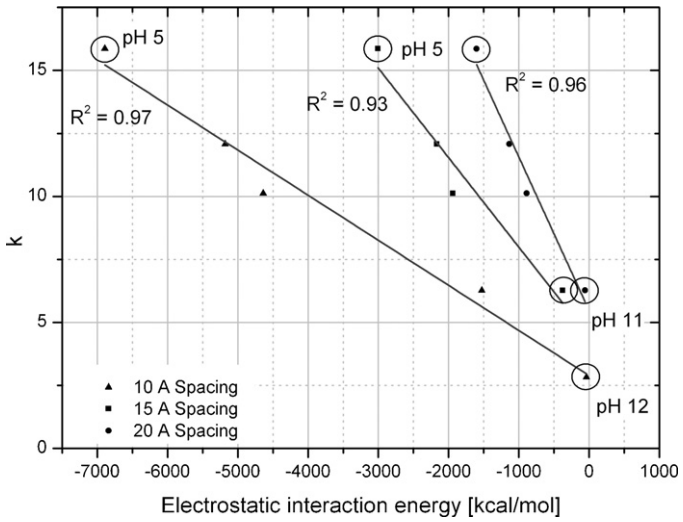


Fig. 11. Correlation of the electrostatic energy for three different ligand spacings and the retention factor $\bar{k}$ calculated from gradient elution studies at different pH.

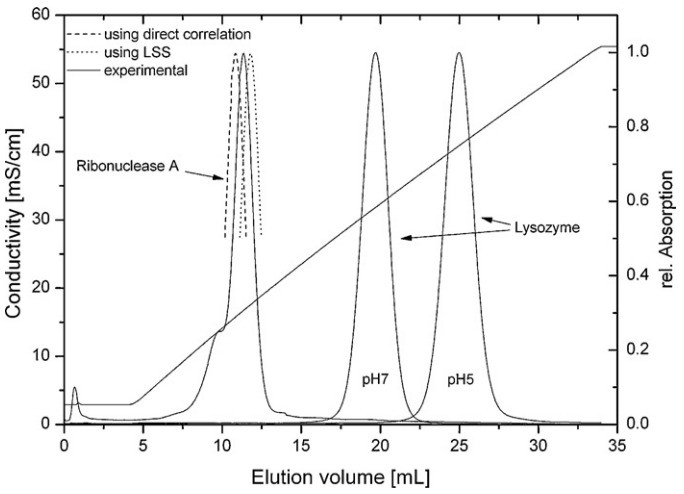


Fig. 12. Comparison between experimental results for the gradient elution of ribonuclease A and lysozyme with the predicted retention behaviour of ribonuclease A using both, the LSS theory and a direct correlation with gradient elution data. The predicted peaks were generated by simply shifting the center of the peak to an elution conductivity according to the predictions.

tion based on the net charge of the protein. For the latter approach, the same correlation was used as described above, except that the net charge was used instead of the average electrostatic interaction energy. The average deviation for the MD based prediction of retention volumes between 1 and 100 CV (corresponding to the data shown in Fig. 10B) was 6% compared to about 450% for the net charge based approach.

3.7. Gradient elution

The linear solvent strength theory (LSS) was developed by Snyder et al. [39] and Dolan et al. [40] for reversed-phase chromatography and extended by Stout et al. for the use in ion-exchange chromatography. According to this theory, gradient elution is not suitable to directly draw information about protein affinities from experiments, as the salt concentration at point to protein elution strongly depends on the slope of the gradient. In contrast to that, isocratic elution studies are under equilibrium conditions and should be used instead. The LSS theory offers a way to calculate average retention factors ($\bar{k}$) from gradient elution data that correlate well with isocratic retention data. It also allows for the calculation of the salt concentration ($\bar{c}$) in the column when the protein peak has migrated half way through the column (for a detailed description please refer to the cited papers), which is then a suitable parameter to describe protein affinities. We used a linear gradient ranging from 0 to 500 mM NaCl over 30 column volumes to elute lysozyme at pHs between 5 and 12. We then used the data to calculate $\bar{k}$ and $\bar{c}$, and correlated these with the average electrostatic interaction energies from the MD simulations. The results for $\bar{k}$ are shown in Fig. 11. The 10 Å ligand spacing showed the best correlation results (R^2 = 0.97).

Interestingly the correlation in Fig. 11 revealed, that at zero electrostatic interaction energy protein retention still occurred, resulting in a k of 2.9. One explanation for the offset is that there were other forces contributing to retention of lysozyme: short-ranged van der Waals forces as well as hydrophobic effects. The van der Waals forces were considerably low in our simulations (in the range of −25 to −60 kcal/mol), did not show an obvious trend and did also not significantly reduce the offset. To calculate reliable energies for the hydrophobic effects, simulations with implicit water molecules would be necessary. These kind of simulations are very cost intensive in terms of computational time and were not performed in this study. A third explanation could be, that the surface used for the simulations was too small to account for all electrostatic interactions, although apparently is was big enough to show a good trend. A larger surface could lead to more negative electrostatic energies at least for the favourable binding orientations which would lower the offset but which would also drastically increase the computational time.

The linear correlation from Fig. 11 was used to calculate $\bar{k}$ and $\bar{c}$ for ribonuclease A: $\bar{k}$ = 6.38 and $\bar{c}$ = 70 mmol/l, resulting in a calculated retention volume of V_r = 11.8 ml. Both parameters were also determined experimentally: $\bar{k}$ = 6.04, $\bar{c}$ = 64 mmol/l and V_r = 11.4 ml (see Fig. 12).

Since there is a linear relationship between the conductivity at point of elution of lysozyme and $\bar{c}$ for $\bar{k} < 2.6$ (according to the LSS theory for the gradient we used), we also directly correlated the conductivity with the electrostatic interaction energy (R^2 = 0.97, correlation not shown) We calculated a conductivity at point of elution of 16.3 mS/cm which referred to an elution volume of V_r = 10.9 ml (see Fig. 12). Both, the LSS based and the direct correlation showed good results with a deviation of the retention volume of about 0.5 ml. Note that the dashed lines representing the predicted retention volumes were generated just by moving the ribonuclease A elution peak to the predicted retention volumes. Shape and width of the peaks were not predicted.

4. Conclusion and outlook

In this paper, a mechanistical model was introduced to describe protein retention based on the calculation of electrostatic energies from molecular mechanics simulations although it should be mentioned, that the relation of the electrostatic interaction energy to the retention volume was made on an empirical basis. The binding of lysozyme to a SP Sepharose FF matrix was simulated at 5 different pH values for 3 different ligand spacings. A surface with a ligand spacing of 10 Å was most suitable to describe the adsorption of lysozyme. A set of possible binding orientations at different pH was in good agreement with experimental data obtained earlier. The electrostatic energy for the interaction between the surface and the protein was found to be strongly dependent on both, the binding orientation and the ligand density of the surface. The average electrostatic energies calculated from 62 different lysozyme orientations were correlated with the conductivity at point of elution for each individual pH obtained by isocratic and linear gradient

elution experiments. The correlation was best for the 10 Å ligand spacing which was also the only ligand spacing that showed negative electrostatic potentials for the interaction with lysozyme at pH 12. All three correlations for the different ligand spacings revealed y-axis intercepts of $\bar{k} > 2.9$ (at zero electrostatic energy) which indicated, that there are other forces contributing to protein retention on SP Sepharose FF and that the electrostatic energies were maybe underestimated due to the limited size of the surface used for the simulations (~100 Å), which will be part of future studies. Nevertheless, the resulting correlation was good enough to successfully predict the retention behaviour of ribonuclease A with a deviation of 6% for isocratic elution and 4.4% (0.5 ml) for gradient elution. Future work aims at the evaluation of the model using a set of other proteins, especially larger ones to show its predictive power. Furthermore it needs to be improved to account for hydrophobic interactions in order to be used more generally (e.g. for other adsorbers, etc.).

References

[1] C.A. Brooks, S.M. Cramer, AIChE J. 38 (12) (1992) 1969.
[2] J.C. Bosma, J.A. Wesselingh, AIChE J. 44 (11) (1998) 2399.
[3] J.C. Bosma, J.A. Wesselingh, AIChE J. 50 (4) (2004) 848.
[4] M.A. Rounds, F.E. Regnier, J. Chromatogr. 283 (1984) 37.
[5] R.R. Drager, F.E. Regnier, J. Chromatogr. 359 (1986) 147.
[6] R.R. Drager, F.E. Regnier, J. Chromatogr. 406 (1987) 237.
[7] M. Vossoughi, I. Alemzadeh, A. Zarrabi, A. Bahari, R. Roostaazad, 3rd WSEAS International Conference on Cellualar and Molecular Biology, Biophysics and Bioengineering, in: C.A. Long, P. Anninos (Eds.), World Scientific and Engineering Acad and Soc, Athens, GREECE, 2007, pp. 7–17.
[8] C. Frerick, P. Kreis, A. Gorak, A. Tappe, D. Melzner, Chem. Eng. Process. 47 (7) (2008) 1128.
[9] T. Vicente, M.F.Q. Sousa, C. Peixoto, J.P.B. Mota, P.M. Alves, M.J.T. Carrondo, J, Membr. Sci. 311 (1–2) (2008) 270.
[10] N. Jakobsson, M. Degerman, E. Stenborg, B. Nilsson, J, Chromatogr. A 1138 (1–2) (2007) 109.
[11] W.D. Chen, H.H. Hu, Y.D. Wang, Chem. Eng. Sci. 61 (21) (2006) 7068.
[12] F. Dismer, M. Petzold, J. Hubbuch, J. Chromatogr. A 1194 (1) (2008) 11.
[13] T. Yang, M.C. Sundling, A.S. Freed, C.M. Breneman, S.M. Cramer, J. Anal. Chem. 79 (23) (2007) 8927.
[14] G. Malmquist, U.H. Nilsson, M. Norrman, U. Skarp, M. Stromgren, E. Carredano, J. Chromatogr. A 1115 (1–2) (2006) 164.
[15] A. Ladiwala, K. Rege, C.M. Breneman, S.M. Cramer, Proc. Natl. Acad. Sci. U.S.A. 102 (33) (2005) 11710.
[16] C.M. Roth, A.M. Lenhoff, Langmuir 9 (4) (1993) 962.
[17] J. Stahlberg, B. Jonsson, C. Horvath, J. Anal. Chem. 63 (17) (1991) 1867.
[18] J. Stahlberg, B. Jonsson, C. Horvath, J. Anal. Chem. 64 (1992) 3118.
[19] J. Stahlberg, J. Chromatogr. A 855 (1) (1999) 3.
[20] G. Raffaini, F. Ganazzoli, Macromol. Biosci. 7 (5) (2007) 552.
[21] H.O. Johansson, J.M. Van Alstine, Langmuir 22 (21) (2006) 8920.
[22] C. Ruggiero, M. Mantelli, A. Curtis, P. Rolfe, Cell Biochem. Biophys. 43 (3) (2005) 407.
[23] P.M. Biesheuvel, M. van der Veen, W. Norde, J. Phys. Chem. B 109 (9) (2005) 4172.
[24] J. Zhou, J. Zheng, S.Y. Jiang, J. Phys. Chem. B 108 (45) (2004) 17418.
[25] F. Dismer, J. Hubbuch, J. Chromatogr. A 1149 (2) (2007) 312.
[26] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G.M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J.M. Wang, P. Kollman, J. Comput. Chem. 24 (16) (2003) 1999.
[27] J.M. Wang, P. Cieplak, P.A. Kollman, J. Comput. Chem. 21 (12) (2000) 1049.
[28] A. Onufriev, D. Bashford, D.A. Case, Proteins: Struct., Funct., Bioinf. 55 (2) (2004) 383.
[29] P. DePhillips, I. Lagerlund, J. Farenmark, A.M. Lenhoff, J. Anal. Chem. 76 (19) (2004) 5816.
[30] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kollman, Comput. Phys. Commun. 91 (1–3) (1995) 1.
[31] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, J. Comput. Chem. 26 (16) (2005) 1668.
[32] W.R. Rypniewski, H.M. Holden, I. Rayment, Biochemistry 32 (37) (1993) 9851.
[33] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (1) (2000) 235.
[34] E. Chatani, R. Hayashi, H. Moriyama, T. Ueki, Protein Sci. 11 (1) (2002) 72.
[35] D. Bashford, Front. Biosci. 9 (2004) 1082.
[36] R.J. Loncharich, B.R. Brooks, R.W. Pastor, Biopolymers 32 (5) (1992) 523.
[37] R.W. Pastor, B.R. Brooks, A. Szabo, Mol. Phys. 65 (6) (1988) 1409.
[38] D.L. Wu, R.R. Walters, J. Chromatogr. 598 (1) (1992) 7.
[39] L.R. Snyder, J.W. Dolan, J.R. Gant, J. Chromatogr. 165 (1) (1979) 3.
[40] J.W. Dolan, J.R. Gant, L.R. Snyder, J. Chromatogr. 165 (1) (1979) 31.