

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

# Coarse-Grained Modeling of The Calcium, Sodium, Magnesium And Potassium Cations Interacting With Proteins

#### Agnieszka Gabriela Lipska (⊠lipska.ag@gmail.com)

University of Gdansk: Uniwersytet Gdanski https://orcid.org/0000-0002-8390-8473

#### Anna Maria Antoniak

University of Gdansk: Uniwersytet Gdanski

#### Patryk Wesołowski

Intercollegiate Faculty of Biotechnology of the University of Gdańsk and the Medical University of Gdańsk: Miedzyuczelniany Wydzial Biotechnologii Uniwersytetu Gdanskiego i Gdanskiego Uniwersytetu Medycznego

#### Alan Warszawski

Intercollegiate Faculty of Biotechnology of the University of Gdańsk and the Medical University of Gdańsk: Miedzyuczelniany Wydzial Biotechnologii Uniwersytetu Gdanskiego i Gdanskiego Uniwersytetu Medycznego

#### Sergey A Samsonov

University of Gdansk: Uniwersytet Gdanski

#### Adam Kazimierz Sieradzan

University of Gdansk: Uniwersytet Gdanski

#### **Research Article**

Keywords: proteins, metal ions, UNRES force field, protein-ion binding

Posted Date: November 16th, 2021

#### DOI: https://doi.org/10.21203/rs.3.rs-1079923/v1

**License:** (a) This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

# Coarse-grained modeling of the calcium, sodium, magnesium and potassium cations interacting with proteins

Agnieszka G. Lipska,<sup>\*,†</sup> Anna M. Antoniak,<sup>†,¶</sup> Patryk Wesołowski,<sup>‡,¶</sup> Alan Warszawski,<sup>‡</sup> Sergey A. Samsonov ,<sup>†</sup> and Adam K.Sieradzan ,<sup>†</sup>

1

2

*+Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland ‡Intercollegiate Faculty of Biotechnology of UG i MUG, Abrahama 58 80-307 Gdańsk, Poland ¶Contributed equally to this work* 

E-mail: lipska.ag@gmail.com

Phone: +48585235350

#### Abstract

3	Metal ions play important biological roles e.g.: activation or deactivation of enzy-
4	matic reactions and signal transduction. Moreover, they can stabilize protein structure,
5	or even be actively involved in the protein folding process. Therefore, accurate treat-
6	ment of the ions is crucial to model and investigate biological phenomena properly.
7	In this work the coarse-grained UNRES (UNited RESidue) force field was extended
8	to include the interactions between proteins and four alkali or alkaline earth metal
9	cations of biological significance, i.e. calcium, magnesium, sodium and potassium. Ad-
10	ditionally, chloride anions were introduced as counter-ions. Parameters were derived
11	from all-atom simulations and incorporate water in an implicit manner. The new force
12	field was tested on the set of the proteins and was able to reproduce the ion-binding
13	preferences.

## 14 Keywords

<sup>15</sup> proteins, metal ions, UNRES force field, protein-ion binding

## <sup>16</sup> Introduction

Approximately 40 % of the protein structures deposited in the Protein Data Bank (PDB)
contain one or more ions.<sup>1</sup> In some cases these ions were introduced artificially to enable
structure determination, nevertheless most of them play important biological roles e.g.:
structure stabilization, activation/deactivation of enzymatic reactions and signal transduction.<sup>2</sup> Therefore, accurate treatment of the ions is crucial to model biological phenomena
properly.

The most abundant ions in proteins structures from the PDB are  $Zn^{2+}$ ,  $Mg^{2+}$  and 23  $Ca^{2+}$ . They have significant effect on the protein folding, stabilize the structure and are an 24 integral part of numerous enzymes.<sup>3</sup> If the cell environment is taken into account  $Na^+$  and 25  $K^+$  are the metal ions with the highest concentration. They take part in signal transduction. 26 Alkali and alkaline earth metal ions have a relatively straightforward pattern of inter-27 actions, in contrast to, for example, transition-metal cations, that form coordinate chemical 28 bonds with protein side-chain atoms. Therefore  $Mg^{2+}$ ,  $Ca^{2+}$ ,  $Na^+$  and  $K^+$  are going to be 29 the main subject of this work. 30

These ions have different interaction preferences for coordinating groups that can be summarized as follows: for  $Ca^{2+}$  Asp, Glu, Asn, Gln and oxygen from the main chain, for  $Mg^{2+}$  Asp, Glu and oxygen from the main chain, and for  $K^+$  and  $Na^+$  oxygen from the main chain. In case of the two last ions the interactions are almost entirely electrostatic but the coordination geometry is essential.<sup>1</sup>

The average concentration of these ions in free state in blood plasma are: 145 mM,<sup>4</sup> 4 mM,<sup>5</sup> 1.5 mM<sup>6</sup> and 1.8 mM<sup>7</sup> for  $Na^+$ ,  $K^+$ ,  $Mg^{2+}$ ,  $Ca^{2+}$  respectively. Disturbed homeostasis of these macroelements can lead to many disorders and diseases, such as: insomnia, anxiety,

urolithiasis or cardiac arrhythmia.<sup>8</sup> Moreover, there is evidence that altered  $Mg^{2+}$  and  $Ca^{2+}$ 39 homeostasis may be related to the progression of neurodegenerative diseases associated 40 with intrinsically disordered proteins (IDP): Alzheimer's disease and Parkinson's disease.<sup>9</sup> 41 While investigating the protein-ion interactions, especially when one aims to obtain 42 information about the dynamics, the use of the methods based on the all-atom force 43 fields seems to be appropriate. The extensive review about metal ion modeling has been 44 published, <sup>10</sup> it covers various strategies in quantum mechanics, classical and polarizable 45 force field methods. Generally, three ion models are frequently used: nonbonded, bonded 46 and with dummy atoms. Those representations are used in many common software 47 like: AMBER,<sup>11–14</sup> CHARMM,<sup>15</sup> and GROMOS.<sup>16</sup> Unfortunately, the use of all-atom (AA) 48 methods require high computational power, especially because most ion-binding proteins 49 are often large multichain subunits. In this regard, coarse-grained (CG) models can be 50 a good alternative. However, introduction of metal ions is a challenging task and this 51 issue is often dismissed in CG representations. In one of the most popular CG force field 52 MARTINI ions are represented as a Q type particle (charged interaction site).<sup>17</sup> In case of 53 single-atom ions the first hydration shell is considered as a part of CG representation. The 54 MARTINI ion model is only qualitatively accurate and the force field itself is unable to 55 model protein folding events.<sup>17,18</sup> 56

Another example is SIRAH force field <sup>19,20</sup> developed to treat proteins with explicit representation of water (WT4) composed by four linked beads and represents 11 tetrahedrallycoordinated water molecules. <sup>21</sup> Hydrated ions:  $Na^+$ ,  $K^+$ ,  $Ca^{2+}$  and  $Cl^-$  are also considered explicitly, each extended ion particle comprising the actual ion and 6 water molecules corresponding to the first hydration shell. <sup>20,21</sup> It has been used in the simulations of natively unfolded proteins and protein aggregation and has the ability to capture ionic strength effects.

<sup>64</sup> The coarse-grained UNited RESidue (UNRES)<sup>22–24</sup> force field was already successfully <sup>65</sup> extended to treat calcium-protein interactions<sup>25</sup> by deriving the interactions of the ion

with Asp, Glu, Asn and Gln side-chains and the peptide group. For the side-chains that do 66 not interact with calcium specifically, simple excluded-volume potentials were introduced. 67 The parameters of the potentials were obtained from potential-energy surfaces of model 68 system:  $Ca^{2+}$ -acetate,  $Ca^{2+}$ -propionate,  $Ca^{2+}$ -acetamide,  $Ca^{2+}$ -propionamide, and  $Ca^{2+}$ -69 N-methylacetamide systems calculated by using ab inito molecular quantum mechanics 70 at the Restricted Hartree-Fock (RHF) level with the 6-31G(d,p) basis set. It should be 71 noted that those calculations were performed *in vacuo* and obtained potentials of mean 72 force (PMFs) were divided by water dielectric constant. Thus, these potentials do not 73 incorporate interactions with solvent and cover only the interactions only with carbonyl 74 groups. 75

In this article we extended the UNRES force field to investigate the interactions between proteins and four alkali and alkaline earth metals of biological significance, i.e. calcium, magnesium, sodium and potassium. Additionally the parameters for chloride ion were also included. The neutrality of the system is not necessary while performing simulations with the UNRES force field but in case when one wants to represent the environment more accurately, especially under high ionic strength, they can be required.

The method of deriving the potentials was modified with regards to the one previously used for the calcium ion.<sup>25</sup> The following two modifications were introduced. First, the new potentials were derived to incorporate interactions with solvent in an implicit manner and therefore the AA force field was used. Second, we considered all amino-acid side-chains and peptide group. Hence the corresponding modifications of the calcium ion-protein interaction center potentials were also recalculated. Introduced potentials have been tested based on the known structures of proteins with bound metal ions.

## **Methods**

#### **90 UNRES force field**

In the coarse-grained UNRES model<sup>22–24</sup> a polypeptide chain is described by two geometry 91 points per residue namely:  $C^{\alpha}$  atom and the geometrical center of each side-chain (SC). 92 Moreover, there are two types of interaction sites: peptide group (p), which is positioned 93 halfway between consecutive  $C^{\alpha}$  atoms and geometrical center of a side-chain. The peptide 94 groups are modeled with spheres whereas side-chains are ellipsoids of revolution along 95 the longer axis (Fig. 1). It should be noted that the  $C^{\alpha}$  atom is not a center of interaction and 96 acts only as a geometric center. The effective energy function is represented by restricted 97 free energy (RFE) or potential of mean force (PMF) of a given conformation ensemble 98 restricted to coarse-grained conformation defined by  $C^{\alpha}$  and SC atoms and is expressed by 99 (eq. 1):<sup>26</sup> 100

$$\begin{aligned} U &= w_{SC} \sum_{i < j} U_{SC_i SC_j} + w_{SCp} \sum_{i \neq j} U_{SC_i p_j} + w_{pp}^{VDW} \sum_{i < j-1} U_{p_i p_j}^{VDW} + w_{pp}^{el} f_2(T) \sum_{i < j-1} U_{p_i p_j}^{el} \\ &+ w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i, \theta_i, \theta_{i+1}) + w_b \sum_i U_b(\theta_i) + w_{rot} \sum_i U_{rot}(\theta_i, \alpha_{SC_i}, \beta_{SC_i}) \\ &+ w_{bond} \sum_i U_{bond}(d_i) + w_{corr}^{(3)} f_3(T) U_{corr}^{(3)} + w_{turn}^{(3)} f_3(T) U_{turn}^{(3)} + U_{ion-ion} + U_{ion-prot} \end{aligned}$$

$$(1)$$

<sup>101</sup> where the *U*'s are energy terms,  $\theta_i$  is the backbone virtual-bond angle between three <sup>102</sup> consecutive C<sup> $\alpha$ </sup> atoms,  $\gamma_i$  is the backbone virtual-bond-dihedral angle between four con-<sup>103</sup> secutive C<sup> $\alpha$ </sup> atoms,  $\alpha_i$  and  $\beta_i$  are the angles defining the location of the center of the united <sup>104</sup> side-chain of residue *i* (Fig. 1) with respect to  $C_{i-1}^{\alpha}$ ,  $C_i^{\alpha}$  and  $C_{i+1}^{\alpha}$  plane,  $d_i$  is the length of <sup>105</sup> the *i*th virtual bond, which is either a C<sup> $\alpha$ </sup> · · · C<sup> $\alpha$ </sup> virtual bond or C<sup> $\alpha$ </sup> · · · SC virtual bond.<sup>24,25</sup> <sup>106</sup> Each energy term is multiplied by an appropriate weight,  $w_x$ , and the terms corre-<sup>107</sup> sponding to factors of order higher than 1 are additionally multiplied by the respective temperature factors that were introduced in paper<sup>22</sup> and which reflect the dependence of the first generalized cumulant term in those factors on temperature, as discussed in refs 22 and 27. The factors  $f_n$  are defined by (eq 2).

$$f_n(T) = \frac{\ln \left[ \exp(1) + \exp(-1) \right]}{\ln \left\{ \exp \left[ (T/T_{\circ})^{n-1} \right] + \exp \left[ - (T/T_{\circ})^{n-1} \right] \right\}}$$
(2)

<sup>111</sup> where  $T_{\circ} = 300$  K.

The term  $U_{SC_iSC_i}$  represents the mean free energy of the hydrophobic (hydrophilic) in-112 teractions between the side chains, which implicitly contains the contributions from the in-113 teractions of the side chain with the solvent. The term  $U_{SC_ip_i}$  denotes the excluded-volume 114 potential of the side-chain – peptide-group interactions. The peptide-group interaction po-115 tential is split into two parts: the Lennard-Jones interaction energy between peptide-group 116 centers  $(U_{p_ip_i}^{VDW})$  and the average electrostatic energy between peptide-group dipoles  $(U_{p_ip_i}^{el})$ ; 117 the second of these terms accounts for the tendency to form backbone hydrogen bonds 118 between peptide groups  $p_i$  and  $p_j$ . The terms  $U_{tor}$ ,  $U_b$ ,  $U_{rot}$ , and  $U_{bond}$  are the virtual-bond-119 dihedral angle torsional terms, virtual-bond angle bending terms, side-chain rotamer, and 120 virtual-bond-deformation terms, respectively; these terms account for the local properties 121 of the polypeptide chain. The term  $U_{corr}^{(3)}$  represent correlation or multibody contributions 122 from the coupling between backbone-local and backbone-electrostatic interactions, and the 123 term  $U_{turn}^{(3)}$  are correlation contributions involving 3 consecutive peptide groups; they are, 124 therefore, termed turn contributions. The term  $U_{ion-ion}$  represents the ion-ion interactions 125 with parameters as in our previous work<sup>25</sup> (charge of each ion was adjusted appropriately). 126 The term  $U_{ion-prot}$  represents the ion-polypeptide-chain interactions (details of obtaining 127 and verifying this term are described in the next parts of this paper). 128

## <sup>129</sup> Determination of the Potentials of Mean Force for protein-ion interac-

130 tions

To obtain the PMFs from which the effective energy terms accounting for protein-ion 131 interactions  $(U_{ion})$  were derived, simulations of the pairs consisting of the ion and the 132 model of the UNRES interaction center (Tab. 1) were performed with the all-atom ff14SB 133 force field from AMBER16 package.<sup>28</sup> Each system was placed in a cuboid box containing 134 TIP3P water molecules,<sup>29</sup> and counter-ions in order to maintain electroneutrality were 135 added if needed. For  $Ca^{2+}$  and  $Mg^{2+}$  the multisite ion model, where the total charges 136 are distributed into n-dummy centers that are placed in the direction of the coordinating 137 atoms (n=7 for  $Ca^{2+}$  and n=6 for  $Mg^{2+}$ ), were used.<sup>30</sup> The simulations of systems with 138  $Na^+$ ,  $K^+$  and  $Cl^-$  were performed with the nonbonded-ions models available in AMBER 139 package.<sup>14</sup> The size of the box was set to ensure minimal distance of 12 Å between any 140 atom in our system and the edge of the box. For each system Umbrella Sampling (US) 141 simulations were performed with different harmonic-restraint potentials imposed on the 142 distance between the centers of the mass of each particle. The distances varied from 2 Å to 143 12 Å, and were sampled every 1 Å. 144

The simulation procedure consisted of four steps: (1) minimization of the starting 145 structure with restrain at 12 Å distance (for 4000 steps of steepest descent and 3500 steps 146 of conjugate gradient); (2) heating of the system with restrain at 12 Å distance until it 147 reached 300 K (NVT, 20 ps, dt=0.002); (3) equalibration of the system for each of the given 148 harmonic restraints value (NPT, 1000 ps, dt=0.002); (4) molecular dynamic of the system at 149 the given harmonic restraints value (NPT, 10000 ps, dt=0.002). To determine the PMFs of 150 the systems studied, we processed the results of all restrained MD simulations by using 151 the weighted histogram analysis method (WHAM)<sup>31</sup> to remove the bias from restraints. 152

# Analytical expressions for the energy of interaction of ions with polypeptide chain

The two types of the UNRES interaction sites cover the 19 amino-acids (as Gly has no side-chain) and the peptide group. They can be divided into three categories, taking into account their physicochemical properties that are responsible for the nature of interaction with the ions: charged interaction sites (Asp, Glu, Arg, Lys), polar interaction sites (peptide group, Ser, Thr, Asn, Gln, Cys, His, Tyr) and hydrophobic sites (Ala, Val, Ile, Leu, Met, Phe, Trp, Pro). Hereby, the following three combinations of analytical expressions for energy interaction of ions with polypeptide chain were used.

For the interactions of the ion with the charged sites, both negatively and positively,
 the following expression were used:

$$E_{charged-ion} = E_{GB} + \Delta F_{cav} + E_{el} + e_{gb} \tag{3}$$

where  $E_{GB}$  is the van der Waals energy term for two ellipsoids expressed as the Gay-Bernetype potential,  $\Delta F_{cav}$  is the cavity term which accounts for the difference between the free energy of hydratation of the side-chain-ion pair and the sum of the free energies of the isolated components,  $E_{el}$  is Coulombic term, and  $e_{gb}$  is generalized Born potential. The expression for the ion-polar sites interactions is defined by:

$$E_{polar-ion} = E_{GB} + \Delta F_{cav} + E_{pol} + E_{cp} \tag{4}$$

where  $E_{pol}$  is the polarization energy, and  $E_{cp}$  is charged-polar term.

<sup>170</sup> For the hydrophobic group the expression is defined by:

$$E_{hydro\,f\,obic-ion} = E_{GB} + \Delta F_{cav} \tag{5}$$

<sup>171</sup> The general analytical expressions with which the effective energy of the interactions

of the ions considered in this study are approximated are shown below.<sup>32–34</sup> The variables
describing the geometry of an UNRES interaction site and an ion are illustrated in Fig. 2.
The Gay-Berne potential<sup>35</sup> is expressed by the equation:

$$E_{GB} = 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}^0}{r_{ij} - \sigma + \sigma_{ij}^0} \right)^{12} - \left( \frac{\sigma_{ij}^0}{r_{ij} - \sigma + \sigma_{ij}^0} \right)^6 \right]$$
(6)

<sup>175</sup> where  $r_{ij}$  is the distance between the center of mass of united side-chain or peptide-group <sup>176</sup> model and ion,  $\varepsilon_{ij}$  is the van der Waals well-depth for a given orientation of the two <sup>177</sup> interaction sites,  $\sigma_{ij}$  is the distance corresponding to the zero value of  $E_{GB}$  for arbitrary <sup>178</sup> orientation of the particles, and  $\sigma_{ij}^0$  is the distance corresponding to the zero value of  $E_{GB}$ <sup>179</sup> for the orientation where  $\theta_{ij} = 90^\circ$ . The dependence of the  $\varepsilon_{ij}$  and  $\sigma_{ij}$  on the orientation of <sup>180</sup> the interacting sites is given by eqs. (7 - 9)

$$\varepsilon_{ij} = \varepsilon_{ij}^{0} \left[ 1 - \chi_{ij}^{\prime(1)} \omega_{ij}^{(1)2} \right]^{2}$$
(7)

$$\sigma_{ij} = \sigma_{ij}^0 \left[ 1 - \chi_{ij}^{(1)} \omega_{ij}^{(1)2} \right]^{-1/2} \tag{8}$$

$$\omega_{ij}^{(1)} = \cos\theta_{ij}^{(1)} \tag{9}$$

where  $\chi_{ij}$  is the van der Waals interaction anisotropy,  $\chi'_{ij}$  is the anisotropy of the van der Waals depth , and  $\varepsilon^0_{ij}$  is the van der Waals depth of the potential energy minimum for the orientation where  $\theta_{ij} = 90^\circ$ . The fitting parameters are:  $\sigma^0_{ij}$ ,  $\varepsilon^0_{ij}$ ,  $\chi_{ij}$ ,  $\chi'_{ij}$ .

<sup>184</sup> To account for the interactions with solvent, a cavity term was introduced:<sup>32,33</sup>

$$\Delta F_{cav} = \frac{\alpha_{ij}^{(1)} \left[ (x \cdot \lambda)^{1/2} + \alpha_{ij}^{(2)} x \cdot \lambda - \alpha_{ij}^{(3)} \right]}{1 + \alpha_{ij}^{(4)} (x \cdot \lambda)^{12}}$$
(10)

185 with

$$x = \frac{r_{ij}}{\sqrt{\sigma_i^2 + \sigma_j^2}} \tag{11}$$

$$\lambda = \left[1 - \chi_{ij}^{\prime\prime(1)} \omega_{ij}^{(1)2}\right]^{1/2}$$
(12)

where  $\alpha_{ij}^{(1)}$ ,  $\alpha_{ij}^{(2)}$ ,  $\alpha_{ij}^{(3)}$ , and  $\alpha_{ij}^{(4)}$  are coefficients,  $\sigma_i^2$  and  $\sigma_j^2$  are the deviation of the Gaussian distribution of the solute density for particles *i* and *j*, respectively, and  $\chi_{ij}^{\prime\prime(1)}$  is the anisotropy pertaining to the cavity term. The fitting parameters are:  $\alpha_{ij}^{(1)}$ ,  $\alpha_{ij}^{(2)}$ ,  $\alpha_{ij}^{(3)}$ ,  $\alpha_{ij}^{(4)}$ ,  $\chi_{ij}^{\prime\prime(1)}$  and  $\sigma = \sqrt{\sigma_i^2 + \sigma_j^2}$ .

The polarization energy originated in the deformation of the charge distribution of the side-chain or peptide group by the ion is expressed by the equation:

$$E_{pol} = 332 \left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}}\right) \left(\frac{\alpha_{ji}^{pol}}{f_{GB}(r'_{ji})}\right)^4$$
(13)

<sup>192</sup> with  $f_{GB}$  being the generalized Born function:<sup>36</sup>

$$f_{GB}(r'_{ji}) = \sqrt{r'_{ji}^2 + a_i a_j exp(-\frac{r'_{ji}^2}{4a_i a_j})}$$
(14)

<sup>193</sup> where the coefficient 322 is there to express the energies in kcal/mol,  $\epsilon_{in}$  and  $\epsilon_{out}$  are the <sup>194</sup> effective dielectric constant of the "inside" of the interacting particles and the solvent (equal <sup>195</sup> to 80 for water), respectively,  $\alpha_{ji}^{pol}$  is the polarizability of the nonpolar part of the particle *i*, <sup>196</sup>  $a_i$  and  $a_j$  are the Born radii of particles *i* and *j*, and  $r'_{ji}$  is the distance between the interacting <sup>197</sup> "head" of a side chain or peptide group and ion. The fitting parameters are:  $a_i$ ,  $\epsilon_{in}$ ,  $\alpha_{ji}^{pol}$  and <sup>198</sup>  $d_{head}$ .

<sup>199</sup> The generalized Born term is expressed by equation:

$$e_{gb} = 332q_i q_j \frac{D}{f_{GB}(r'_{ji})}$$
(15)

200 with

$$D = \frac{1}{\epsilon_{in}} - \frac{exp(-\kappa f_{GB}(r'_{ji}))}{\epsilon_{out}}$$
(16)

where  $\kappa$  is Debay screening length. The fitting parameters are:  $\kappa$  and  $a_i$ ,  $\epsilon_{in}$  and  $d_{head}$ .

<sup>202</sup> The Coulombic term is expressed by the equation:

$$E_{el} = 332 \left(\frac{q_1 q_2}{\epsilon_{in} r'_{ij}}\right) \tag{17}$$

where  $q_1$  and  $q_2$  are the charges of particles *i* and *j*, respectively. The fitting parameters are:  $\epsilon_{in}$  and  $d_{head}$ .

<sup>205</sup> The charged-polar term is expressed by equation:<sup>34</sup>

$$E_{cp} = w_{\perp} \cdot \frac{q \cdot \cos \theta_{ij}}{r_{ij}^{\prime 2}} - w_{\parallel} \cdot \frac{q^2 \cdot \sin \theta_{ij}^2}{r_{ij}^{\prime 4}}$$
(18)

where  $w_{\perp}$  and  $w_{\parallel}$  are fitting parameters, and  $r'_{ij}$  is a distance between the ion and the center of the amphiphilic head group of side-chain or peptide-group. The fitting parameters are:  $w_{\perp}, w_{\parallel}$  and  $d_{head}$ .

The approximate analytical expressions for ion-interaction-site energy functions were fitted to the PMF calculated numerically for each ion-interaction-site system by means of a nonlinear least-squares Marquardt method. The similar procedure was used before in introducing the potentials for protein-DNA interaction into UNRES force field<sup>34</sup> Subsequently, the obtained expressions were compatibly added to the already existing UNRES energy function.

#### Verification of the protein-ion potentials 215

#### Simulation procedures 216

The developed coarse-grained model of polypeptide-ion interaction was tested to predict 217 the ion-binding sites on five proteins listed in Tab. 2. 218

A set of multiplexed replica exchange simulations (MREMD)<sup>37</sup> were performed in 219 temperatures 260 K, 272 K, 279 K, 284 K, 288 K, 291 K, 294 K, 298 K, 308 K, 322 K, 341 K, 220 370 K. The temperature set was adjusted with the algorithm developed by Trebst et al.<sup>38</sup> 221 which maximizes the walk of replicas in temperature. For each temperature 4 replicas 222 were calculated. Every simulation took 10 mln steps with a time step equal to 4.89 fs. The 223 periodic boundary conditions<sup>39</sup> were used with the cubic boxes edges length equal to 224 150 Å. The simulations were started from the experimental structures from the Protein 225 Data Base (PDB) - including the position of ions. This starting point is sufficient because 226 the MREMD allows to enhance probing of the conformation space and ions are able to 227 dissociate and bound again to the protein. For analysis, only the second half of trajectories 228 were considered. During the simulation temperatures were exchanged between trajectories 229 every 10000 steps, therefore to obtain the correct ensemble properties reweighing was 230 performed with the bin-less Weighted Histogram Analysis Method (WHAM), which was 231 implemented in UNRES and described in earlier work.<sup>22</sup> Subsequently, the five most 232 populated clusters were obtained with the use of Ward's minimum variance method<sup>40</sup> in 233 the temperature 280 K. 234

The simulations in higher ion concentrations were performed for the systems where 235 ions dissociate permanently after starting from the PDB structure. It was the case for two 236 proteins: 2Z2K and 5E56, and used concentrations were 55 mM and 70 mM respectively. 237 For this test counter-ions were used but the rest of the procedure was the same as the one 238 described above. 230

In the last test association binding constants of  $Ca^{2+}$ ,  $Mg^{2+}$ ,  $Na^+$  and  $K^+$  to the  $\alpha$ -

lactalbumin were computed and compared to experimental values. For this purpose the molecular dynamic simulations were performed for the 1F6S<sup>41</sup> protein with the ion ( $Ca^{2+}$ ,  $Mg^{2+}$ ,  $Na^+$  or  $K^+$ ) bound in the native position. The 200 independent trajectories were run at 293 K, each last for 5 mln steps with time step equal to 4.89 fs. The periodic boundary conditions were the same as in two other tests.

#### 246 Analysis

For simulations starting from the experimental structure the following analysis was performed. First, the representative structure for each cluster (the conformation with the lowest RMSD to cluster average structure) was determined and the TM-score was calculated in order to check its quality.

<sup>251</sup> Subsequently, the distances averaged over all structures for ion from its starting place-<sup>252</sup> ment for every cluster were calculated. Normalized histograms of those distances were <sup>253</sup> also prepared.

<sup>254</sup> Moreover, the contact maps for ions and particular amino acid side-chains averaged <sup>255</sup> over all structures were prepared for each cluster and compared with the corresponding <sup>256</sup> experimental structure. The scoring function ( $S_{i,j}$ ) for the contact between any  $C^{\alpha}$  atoms <sup>257</sup> and ion ( $d_{i,j}$ ) is defined as follows:<sup>42</sup>

$$S_{i,j} = \begin{cases} 0 & \text{for } d_{i,j} > 12\\ \frac{12 - d_{i,j}}{4} & \text{for } 8 \le d_{i,j} \le 12\\ 1 & \text{for } d_{i,j} < 8 \end{cases}$$
(19)

<sup>258</sup> Root-mean square fluctuation (RMSF) was used to analyze the fluctuation of  $C^{\alpha}$  atoms <sup>259</sup> in each cluster. As a reference structure the cluster representative was taken.

For simulations with higher ion concentration the cluster representatives were also verified by TM-score. Then the radial distribution function (RDF) for the distances between <sup>262</sup> ions and protein were calculated.

In order to calculate the binding constants the concentration of unbound states as a function of time averaged over all 200 trajectories were calculated for each system. An unbound state was defined as a structure where ions were further than 8 Å from any  $C^{\alpha}$ atom of the protein. Subsequently, function expressed by eq. 20 was fitted with the use of the nonlinear least-squares (NLLS) Marquardt-Levenberg algorithm<sup>43</sup> implemented in the GNUPLOT program (http://www.gnuplot.info).

$$[E](t) = 2[E]_0 k_{-1} \left( \frac{1 - \exp(-tW)}{k_{-1}(1 - \exp(-tW)) + W(1 + \exp(-tW))} \right)$$
(20)

where E[t] is the concentration of unbound ion and W is expressed by:

$$W = \sqrt{k_{-1}^2 + 4k_1k_{-1}[E]_0} \tag{21}$$

The eq. 20 is derived based on the method used in.<sup>44</sup>

### 271 Results and discussion

#### 272 Obtained Potentials of Mean Force and fitted functions

The PMFs, fitted analytical energy functions and parameters are presented in SI (Fig. SI 273 1-15 and Tab SI 1-3). The strongest interactions appeared for the systems containing ions: 274  $Ca^{2+}$  and  $Mg^{2+}$ , and side-chain models for: Asp, Glu (Fig. SI 3A-D and Fig. SI 6A-D). 275 This is reasonable as there is the strongest Coulombic interaction for oppositely charged 276 molecules. Those two ions strongly interact also with Asn, Gln and the peptide-group 277 models (Fig. SI 2 G-J,O-P and Fig. SI 5 G-J,O-P). For  $K^+$  and  $Na^+$  the strongest interaction 278 is observed for Asp and Glu side-chain models (Fig. SI 9 A-D and Fig. SI 12 A-D). As 279 in case of  $Ca^{2+}$  and  $Mg^{2+}$  the strongest interactions are observed for oppositely charged 280

molecules. The next strongest interactions are for the peptide-group model (Fig. SI 8 O-P 281 and Fig. SI 11 O-P), Asn and Gln models (Fig. SI 8 G-J and Fig. SI 11 G-J) and weaker for 282 His model (Fig. SI 8 K-L and Fig. SI 11 K-L). For all metal ions also weak interactions for 283 Ser model (Fig. SI 2 A-B, Fig. SI 5 A-B, Fig. SI 8 A-B, Fig. SI 11 A-B) and Thr model (Fig. SI 284 2 C-D, Fig. SI 5 C-D, Fig. SI 8 C-D, and Fig. SI 11 C-D) were observed. Those patterns are 285 mostly in agreement with the information collected from the PDB analysis.<sup>1</sup> As for  $Cl^{-}$ 286 the interactions are observed for Arg and Ser models only (Fig. SI G-H and Fig. SI 14 A-B) 287 but because of their weak character they are represented by even weaker CG potentials. 288

#### 289 Simulations on the test proteins set

Probabilities and population sizes of each cluster, TM-score for cluster representative and average ion displacement are presented in Tab. 3. The probabilities and population sizes of each cluster represent satisfactory distribution of obtained structures. The TM-score shows that at least one cluster in each system has in about the same fold (the value is greater than  $\sim 0.5$ ).

Normalized histograms of distances between the predicted ion placement and the one in the native structure are in Fig. 3. As assumed, even for the metal ions that strongly interact with the protein, the ion was able to dissociate from the molecule during the simulation (panels A and B). Whereas for the rest of the ions the broader distances are represented (panels C, D and E) indicating weaker interactions.

The contact maps for ion-protein interactions (Fig. 4) and structures (Fig. 5) are presented for proteins 6EKG, 4AQO and 4RJ9 because only for those systems contacts occur (Fig. 3).

The best prediction of ion placement was determined for 4AQO protein with  $Ca^{2+}$ , especially for the third and fifth clusters with average ion displacements 6.52 Å and 2.99 Å , respectively (Tab. 3). The quality of predicted binding sites can be verified based on contact maps (Fig. 4) where the prediction of the amino-acid side-chains that interact with  $Ca^{2+}$  is in very good agreement with the experimental data. Moreover, the structure of the representative of the cluster with the best prediction (cluster III, based on contact maps) compared with the experimental structure is presented in Fig. 5.

 $Mg^{2+}$ - binding protein (6EKG) strongly interacts with ion (Fig. 3) but based on average ion displacements (Tab. 3) and contact maps (Fig. 4) the predicted binding sites are different from the experimental one. The best prediction is the structure form cluster IV and is presented in Fig. 5.

High-quality prediction was also obtained for  $K^+$  (4RJ9) but only in the least populated cluster. For this protein the maps for two  $K^+$  were shown in Fig. 4 where weak but correct contacts can be observed. The best prediction (the structure form cluster V) is presented in Fig. 5.

The worst predictions were obtained for the interactions with  $Na^+$  and  $Cl^-$  (proteins 5E56 and 2Z2K, respectively), where ion dissociate from the protein and do not bind to the protein again (Fig. 3 D and E).

RMSF plots are presented in SI (Fig. SI 16). Shaded areas indicate the amino acid residues that are in contact with the ion in the native structure. The lowest fluctuations are observed for 5E56 protein (Fig SI 16 D) meaning the structures in each cluster are the most similar to each other. For the rest of the systems fluctuations are higher but still showing the similarities in each cluster. Only for 4AQO protein the highest fluctuations can be observed for the region where ion bounds.

The results from simulations in high ion concentration are presented in Fig 6 A. It can be seen that neither  $Cl^-$  nor  $Na^+$  interact with protein. For comparison RDF for the simulations that were started from the experimental structures (low ion concentrations) were calculated for  $Ca^{2+}$  and  $K^+$  (Fig 6 B). The strong interactions with both ions can be observed.

<sup>332</sup> The plots of concentration of unbound states of bovine  $\alpha$ -lactalbumin as a function <sup>333</sup> of time and curves (eq. 20) fitted to those data are in Fig SI 16. The calculated and experimental equilibrium binding constant are in Tab 4. Despite the fact that qualitatively binding affinities are not reproduced the tendency of binding affinity is in agreement for the simulations and experiment with the strongest binding for  $Ca^{2+}$  then  $Mg^{2+}$ ,  $Na^+$  and  $K^+$ , respectively.

## **338 Conclusions**

The new approach to simulate the protein-ion interactions in a CG representation was reported in this work. We implemented  $Ca^{2+}$ ,  $Mg^{2+}$ ,  $Na^+$ ,  $K^+$  and  $Cl^-$  within the formalism of the physics-based UNRES model. Parameters were derived from AA simulations and incorporate water in an implicit manner. It is important to remember that the results can be at best of the same quality as the ones obtained by the method used to derive the new potentials.

<sup>345</sup> We were able to predict the binding site of  $Ca^{2+}$ ,  $Mg^{2+}$  and  $K^+$  and the strongest <sup>346</sup> interaction were observed for  $Ca^{2+}$  and  $Mg^{2+}$ . As for the simulations of  $Na^+$  and  $Cl^-$  no <sup>347</sup> interactions with protein were observed. Based on the PDB analysis<sup>1</sup>  $Na^+$  (and to some <sup>348</sup> extend  $K^+$ ) with water molecules around often appear in the bulk solvent in the crystal <sup>349</sup> structure, or with a small number of protein oxygen atoms around them. The interactions <sup>350</sup> are likely to be very labile and would not be expected to persist in the solution,<sup>1</sup> and our <sup>351</sup> simulations reflect that.

<sup>352</sup> Moreover, we were able to reproduce the experimental tendency of binding of the <sup>353</sup> metal ions to bovine  $\alpha$ -lactalbumin that again verified the strongest affinity for  $Ca^{2+}$  and <sup>354</sup>  $Mg^{2+}$ .

Therefore, the new potentials allow to represent approximate interaction sites for  $Ca^{2+}$ ,  $Mg^{2+}$  and  $K^+$ . They also can be applied to reproduce the labile interaction pattern for  $Na^+$ and  $K^+$ .

## **Supplementary information**

The supplementary information contains the PMFs, fitted analytical energy function, parameters, plots of RMSF of  $C^{\alpha}$  atoms in each cluster for protein test set and plots of concentration of unbound structures of 1F6S and metal ions over time with fitted kinetic curves.

## **J63 Declarations**

### <sup>364</sup> Competing Interests:

<sup>365</sup> The authors declare no competing interests

#### 366 Funding

<sup>367</sup> The funding for this work was obtained from the National Science Centre (Poland): UMO-

<sup>368</sup> 2016/21/N/ST4/03154 (to AGL).

#### <sup>369</sup> Availability of data and material

The datasets generated during and analysed during the current study are available from the corresponding author on reasonable request.

#### **Code availability**

<sup>373</sup> The code is available through GIT: git://mmka.chem.univ.gda.pl/unres4 as UCGM branch

#### 374 Authors' contributions

<sup>375</sup> AGL gave the main idea, performed the AA simulation, fitted the analytical expressions for <sup>376</sup>  $K^+$ , performed and analyzed the test of the method, and prepared the manuscript. AMA fitted the analytical expressions for  $Cl^-$  and  $Na^+$ . PW fitted the analytical expressions for  $Ca^{2+}$  an helped with the Introduction section of the manuscript. AW fitted the analytical expressions for  $Mg^{2+}$ . SAS helped with the AA simulations. AKS helped with implantation into UNRES software, performed the analysis for kinetic constants and helped with the rest of data analysis. All authors read and revised the final version of the manuscript.

## **382** Acknowledgement

The funding for this work was obtained from the National Science Centre (Poland): UMO-2016/21/N/ST4/03154 (AGL). Computational resources were provided by (a) the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk, and (b) the Interdisciplinary Center for Mathematical and Computer Modeling in Warsaw (ICM), grant GA76-18 (c) our 682-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk.

## **References**

- (1) Harding, M. M.; Nowicki, M. W.; Walkinshaw, M. D. Metals in protein struc tures: a review of their principal features. *Crystallogr. Rev.* 2010, *16*, 247–302, DOI:
   10.1080/0889311X.2010.485616.
- (2) Rosenzweig, A. C. Metallochaperones: bind and deliver. *Chem. Biol.* 2002, *9*, 673–677,
   DOI: 10.1016/S1074-5521(02)00156-4.
- (3) Dudev, T.; Lim, C. Principles governing Mg, Ca, and Zn binding and selectivity in
   proteins. *Chem. Rev.* 2003, 103, 773–788, DOI: 10.1021/cr020467n.
- <sup>397</sup> (4) Kumar, S.; Berl, T. Sodium. *The Lancet* **1998**, *352*, 220–228, DOI: <sup>398</sup> 10.1016/S0140-6736(97)12169-9.

- <sup>399</sup> (5) Pratt, P. F. Potassium. *Methods of Soil Analysis: Part 2 Chemical and Microbiological* <sup>400</sup> Properties 1965, 9, 1022–1030, DOI: 10.2134/agronmonogr9.2.c20.
- (6) Jahnen-Dechent, W.; Ketteler, M. Magnesium basics. *Clin. Kidney J.* 2012, *5*, i3–i14, DOI:
   10.1093/ndtplus/sfr163.
- 403 (7) Bushinsky, D. A.; Monk, R. D. Calcium. *The Lancet* 1998, 352, 306–311, DOI:
   404 10.1016/S0140-6736(97)12331-5.
- (8) Sigel, A.; Sigel, H.; Sigel, R. K. Interrelations between essential metal ions and human
   diseases; Springer, 2013; Vol. 13.
- (9) Breydo, L.; Uversky, V. N. Role of metal ions in aggregation of intrinsically disor dered proteins in neurodegenerative diseases. *Metallomics* 2011, *3*, 1163–1180, DOI:
   10.1039/c1mt00106j.
- (10) Li, P.; Merz Jr, K. M. Metal ion modeling using classical mechanics. *Chem. Rev.* 2017,
   117, 1564–1686, DOI: 10.1021/acs.chemrev.6b00440.
- (11) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz Jr, K. M. Rational design of particle
  mesh Ewald compatible Lennard-Jones parameters for+ 2 metal cations in explicit
  solvent. *J. Chem. Theory Comput.* 2013, *9*, 2733–2748, DOI: 10.1021/ct400146w.
- (12) Li, P.; Merz Jr, K. M. Taking into account the ion-induced dipole interaction in
  the nonbonded model of ions. *J. Chem. Theory Comput.* 2014, 10, 289–297, DOI:
  10.1021/ct400751u.
- (13) Li, P.; Song, L. F.; Merz Jr, K. M. Parameterization of highly charged metal ions
  using the 12-6-4 LJ-type nonbonded model in explicit water. *J. Phys. Chem. B* 2015, *119*,
  883–895, DOI: 10.1021/jp505875v.
- (14) Li, P.; Song, L. F.; Merz Jr, K. M. Systematic parameterization of monovalent ions

- employing the nonbonded model. *J. Chem. Theory Comput.* 2015, *11*, 1645–1657, DOI:
   10.1021/ct500918t.
- (15) MacKerell Jr, A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr, R. L.; Evanseck, J. D.;
  Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., et al. All-atom empirical potential
  for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 1998, 102,
  3586–3616, DOI: 10.1021/jp973084f.
- (16) Scott, W. R.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. The GROMOS
  biomolecular simulation program package. *J. Phys. Chem. A* 1999, *103*, 3596–3607, DOI: 10.1021/jp984217f.
- (17) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The
  MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* 2007, *111*, 7812–7824, DOI: 10.1021/jp071097f.
- (18) Marrink, S. J.; Corradi, V.; Souza, P. C.; Ingolfsson, H. I.; Tieleman, D. P.; Sansom, M. S.
  Computational modeling of realistic cell membranes. *Chem. Rev.* 2019, 119, 6184–6226,
  DOI: 10.1021/acs.chemrev.8b00460.
- (19) Darré, L.; Machado, M. R.; Brandner, A. F.; González, H. C.; Ferreira, S.; Pantano, S.
  SIRAH: a structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J. Chem. Theory Comput.* 2015, *11*, 723–739, DOI: 10.1021/ct5007746.
- (20) Machado, M. R.; Barrera, E. E.; Klein, F.; Sóñora, M.; Silva, S.; Pantano, S. The SIRAH
  2.0 force field: altius, fortius, citius. *J. Chem. Theory Comput.* 2019, *15*, 2719–2733, DOI:
  10.1021/acs.jctc.9b00006.
- (21) Darré, L.; Machado, M. R.; Dans, P. D.; Herrera, F. E.; Pantano, S. Another coarse grain

model for aqueous solvation: WAT FOUR? *J. Chem. Theory Comput.* 2010, *6*, 3793–3807,
 DOI: 10.1021/ct100379f.

(22) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Ołdziej, S.; Wachucik, K.;
Scheraga, H. A. Modification and optimization of the united-residue (UNRES) potential
energy function for canonical simulations. I. Temperature dependence of the effective
energy function and tests of the optimization method with single training proteins. *J. Phys. Chem. B* 2007, *111*, 260–285, DOI: 10.1021/jp065380a.

- (23) Voth, G. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*, 1st ed.; Voth, G.,
   Ed.; CRC Press, Taylor & Francis Group, 2008; Chapter 1, pp 1–4.
- (24) Liwo, A. et al. A unified coarse-grained model of biological macromolecules based
  on mean-field multipole-multipole interactions. *J. Mol. Model.* 2014, 20, 2306, DOI:
  10.1007/s00894-014-2306-5.
- (25) Khalili, M.; Saunders, J. A.; Liwo, A.; Ołdziej, S.; Scheraga, H. A. A united residue
  force-field for calcium–protein interactions. *Protein Sci.* 2004, *13*, 2725–2735, DOI:
  10.1110/ps.04878904.

(26) Liwo, A.; Sieradzan, A. K.; Lipska, A. G.; Czaplewski, C.; Joung, I.; Żmudzińska, W.;
Hałabis, A.; Ołdziej, S. A general method for the derivation of the functional forms of
the effective energy terms in coarse-grained energy functions of polymers. III. Determination of scale-consistent backbone-local and correlation potentials in the UNRES force
field and force-field calibration and validation. *J. Chem. Phys.* 2019, 150, 155104, DOI:
10.1063/1.5093015.

(27) Shen, H.; Liwo, A.; Scheraga, H. A. An improved functional form for the temperature
scaling factors of the components of the mesoscopic UNRES force field for simulations of protein structure and dynamics. *J. Phys. Chem. B* 2009, *113*, 8738–8744, DOI:
10.1021/jp901788q.

(28) Case, D. A.; Betz, R.; Cerutti, D.; Cheatham Iii, T.; Darden, T.; Duke, R.; Giese, T.;
Gohlke, H.; Goetz, A.; Homeyer, N., et al. AMBER 2016. *University of California, San Francisco* 2016, *810*.

(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 1983, 79,
926–935, DOI: 10.1063/1.445869.

(30) Saxena, A.; Sept, D. Multisite ion models that improve coordination and free energy
calculations in molecular dynamics simulations. *J. Chem. Theory Comput.* 2013, *9*, 3538–
3542, DOI: 10.1021/ct400177g.

(31) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The
weighted histogram analysis method for free-energy calculations on biomolecules. I.
The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021, DOI: 10.1002/jcc.540130812.

(32) Makowski, M.; Liwo, A.; Maksimiak, K.; Makowska, J.; Scheraga, H. A. Simple
physics-based analytical formulas for the potentials of mean force for the interaction of
amino acid side chains in water. 2. Tests with simple spherical systems. *J. Phys. Chem. B*2007, 111, 2917–2924, DOI: 10.1021/jp065917k.

(33) Makowski, M.; Sobolewski, E.; Czaplewski, C.; Liwo, A.; Ołdziej, S.; No, J. H.; Scheraga, H. A. Simple Physics-Based Analytical Formulas for the Potentials of Mean Force for the Interaction of Amino Acid Side Chains in Water. 3. Calculation and Parameterization of the Potentials of Mean Force of Pairs of Identical Hydrophobic Side Chains. *J. Phys. Chem. B* 2007, *111*, 2925–2931, DOI: 10.1021/jp065918c.

(34) Yin, Y.; Sieradzan, A. K.; Liwo, A.; He, Y.; Scheraga, H. A. Physics-based potentials for
 coarse-grained modeling of protein–DNA interactions. *J. Chem. Theory Comput.* 2015,
 11, 1792–1808, DOI: 10.1021/ct5009558.

- (35) Gay, J. G.; Berne, B. J. Modification of the overlap potential to mimic a linear site-site
  potential. *J. Chem. Phys.* **1981**, *74*, 3316–3319, DOI: 10.1063/1.441483.
- (36) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of
   solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 1990, 112, 6127–6129,
   DOI: 10.1021/ja00172a038.
- <sup>500</sup> (37) Czaplewski, C.; Kalinowski, S.; Liwo, A.; Scheraga, H. A. Application of multiplexing <sup>501</sup> replica exchange molecular dynamics method to the UNRES force field: tests with  $\alpha$ <sup>502</sup> and  $\alpha + \beta$  proteins. *J. Chem. Theory Comput.* **2009**, *5*, 627–640, DOI: 10.1021/ct800397z.
- (38) Trebst, S.; Troyer, M.; Hansmann, U. H. Optimized parallel tempering simulations of
   proteins. J. Chem. Phys. 2006, 124, 174903, DOI: 10.1063/1.2186639.
- (39) Sieradzan, A. K. Introduction of periodic boundary conditions into UNRES force field.
   *J. Comput. Chem.* 2015, *36*, 940–946, DOI: 10.1002/jcc.23864.
- <sup>507</sup> (40) Späth, H. Cluster Analysis Algorithms; Halsted Press; New York, 1980.
- (41) Chrysina, E. D.; Brew, K.; Acharya, K. R. Crystal structures of apo-and holo-bovine
   *α*-lactalbumin at 2.2-Å resolution reveal an effect of calcium on inter-lobe interactions.
   *J. Biol. Chem.* 2000, 275, 37021–37029, DOI: 10.1074/jbc.M004752200.
- (42) Peng, X.; Sieradzan, A. K.; Niemi, A. J. Thermal unfolding of myoglobin
  in the Landau-Ginzburg-Wilson approach. *Phys. Rev. E* 2016, *94*, 062405, DOI:
  10.1103/PhysRevE.94.062405.
- (43) Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters.
   *J. Soc. Ind. Appl. Math* 1963, *11*, 431–441, DOI: 10.1137/0111030.
- <sup>516</sup> (44) Spodzieja, M.; Kuncewicz, K.; Sieradzan, A.; Karczyńska, A.; Iwaszkiewicz, J.; Ces-
- son, V.; Węgrzyn, K.; Zhukov, I.; Maszota-Zieleniak, M.; Michielin, O., et al. Disulfide-

- linked peptides for blocking BTLA/HVEM binding. Int. J. Mol. Sci. 2020, 21, 636, DOI:
   10.3390/ijms21020636.
- (45) Eckhard, U.; Schönauer, E.; Brandstetter, H. Structural basis for activity regulation
   and substrate preference of clostridial collagenases G, H, and T. J. Biol. Chem. 2013, 288,
   20184–20194, DOI: 10.1074/jbc.M112.448548.
- (46) Quax, T. E.; Altegoer, F.; Rossi, F.; Li, Z.; Rodriguez-Franco, M.; Kraus, F.; Bange, G.;
   Albers, S.-V. Structure and function of the archaeal response regulator CheY. *Proc. Natl. Acad. Sci. U.S.A.* 2018, *115*, E1259–E1268, DOI: 10.1073/pnas.1716661115.
- (47) Yung, Y.-L.; Cheung, M.-Y.; Miao, R.; Fong, Y.-H.; Li, K.-P.; Yu, M.-H.; Chye, M.-L.;
  Wong, K.-B.; Lam, H.-M. Site-directed mutagenesis shows the significance of interactions with phospholipids and the G-protein OsYchF1 for the physiological functions of
  the rice GTPase-activating protein 1 (OsGAP1). *J. Biol. Chem.* 2015, 290, 23984–23996,
  DOI: 10.1074/jbc.M115.655639.
- (48) Marshall, J. D.; Heeke, D. S.; Rao, E.; Maynard, S. K.; Hornigold, D.; McCrae, C.;
  Fraser, N.; Tovchigrechko, A.; Yu, L.; Williams, N., et al. A novel class of small molecule
  agonists with preference for human over mouse TLR4 activation. *PLoS One* 2016, *11*,
  e0164632, DOI: 10.1371/journal.pone.0164632.
- (49) Selvaraj, M.; Roy, S.; Singh, N.; Sangeetha, R.; Varshney, U.; Vijayan, M. Structural
  plasticity and enzyme action: crystal structures of Mycobacterium tuberculosis peptidyltRNA hydrolase. *J. Mol. Biol.* 2007, 372, 186–193, DOI: 10.1016/j.jmb.2007.06.053.
- (50) Putignano, V.; Rosato, A.; Banci, L.; Andreini, C. MetalPDB in 2018: a database of
   metal sites in biological macromolecular structures. *Nucleic Acids Res.* 2018, 46, D459–
   D464, DOI: 10.1093/nar/gkx989.
- 541 (51) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.;
  - 25

- Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* 2000, 28, 235–242, DOI: 10.1093/nar/28.1.235.
- 544 (52) Permyakov, E. A.; Morozova, L. A.; Burstein, E. A. Cation binding effects on the pH,
- thermal and urea denaturation transitions in  $\alpha$ -lactalbumin. *Biophys. Chem.* **1985**, 21,
- <sup>546</sup> 21–31, DOI: 10.1016/0301-4622(85)85003-1.

## 547 Figure legends

**Fig. 1.** The UNRES model of polypeptide chains. The interaction sites are peptide-group 548 centers (p), and side-chain centers (SC) attached to the corresponding  $\alpha$ -carbons. The 549 geometry of the chain can be described by virtual-bond lengths, backbone virtual-550 bond angles  $\theta_i$ , backbone virtual-bond-dihedral angles  $\gamma_i$ , and the angles  $\alpha_i$  and  $\beta_i$ 551 that describe the location of a side chain with respect to the coordinate frame defined 552 by  $C_{i-1}^{\alpha}$ ,  $C_i^{\alpha}$  and  $C_i^{\alpha}$ ,  $C_{i+1}^{\alpha}$ . Reprinted with permission from Zaborowski, Bartłomiej, et 553 al. "A maximum-likelihood approach to force-field calibration." J. Chem. Inf. Model. 554 55.9 (2015): 2050-2070. Copyright 2015 American Chemical Society. 555

Fig. 2. The model of the interaction of the UNRES interaction site (particle i) and ion (particle j) where:  $d_{head}$  is the distance from the center of particle i and the interacting head of it,  $r_{ij}$  is the distance between the center of the particle i and particle j,  $\hat{r}_{ij}$ is the unit vector along the  $r_{ij}$  distance,  $\hat{u}_{ij}$  is the unit vector along the long axis of particle i,  $\theta_{ij}$  is the angle between vector  $u_{ij}$  and vector  $r_{ij}$ ,  $r'_{ji}$  is the distance between the interacting "head" of a side chain and ion.

Fig. 3. Normalized histograms of distances between the predicted ion placement and the
 one in the native structure for: (A) 4AQO, (B) 6EKG, (C) 4RJ9, (D) 5E56 and (E) 2Z2K.
 Results from the simulations started from the experimental structure.

Fig. 4. Contact maps for ions and particular amino acid side-chains averaged over all
 structures in clusters in comparison with the contact maps for the native structure.

Fig. 5. The experimental structures (cartoon representation, grey) and the predicted coarse grained models (sicks representation, red) for three out of five tested proteins. Ions
 are represented as spheres. The predicted models are the representatives of the
 best cluster. The distances between predicted ion placement and the one in the
 native structure are in the brackets. Results from the simulations started from the

<sup>572</sup> experimental structure.

<sup>573</sup> Fig. 6. RDF values (A) obtained from simulations under high ionic strength for interactions

between  $Cl^-$  (2Z2k, blue line) and  $Na^+$  (5E56, red line); and (B) from simulations

under low ionic strength for interactions between  $Ca^{2+}$  (4AQO, blue line) and  $K^+$ 

<sup>576</sup> (4RJ9, red line).

Interaction site	model
peptide group	N-methylacetamide
Ala	ethane
Pro	n-propane
Val	isobutene
Leu/Ile	isopentane
Phe	ethylbenzene
Met	methyl propyl sulfide
Trp	indole
Asp	propionate anion
Glu	butyrate anion
Lys	n-pentylamine cation
Arg	1-butylgaunidine cation
Ser	etanol
Tyr	4-ethylphenol
Thr	2-propanol
His	5-ethylimidazole
Cys	ethyl sulfide
Asn	ethanamide
Gln	proanamide

Table 1: Models of UNRES interaction centers.

PDB ID	sequence length	ions	amino-acid side-chains interacting with ion	ref
4AQO	86	$Ca^{2+}$	Asn1, Lys2, Asp29, Asp31, Asp70 <sup><i>a</i></sup>	45
6EKG	121	$Mg^{2+}$	Asp10, Asp55, Val57 <sup><i>a</i></sup>	46
4RJ9	161	$K^{+}, K^{+}$	1: Arg18, Asp77, 2: Asp24, Asp70, Lys71 <sup><i>a</i></sup>	47
5E56	116	$Na^+$	Phe60, Tyr63, Cys66 <sup>a</sup>	48
2Z2K	180	$Cl^{-}$	His43, Lys44, Arg45 <sup>b</sup>	49

Table 2: Test set proteins

<sup>577</sup> <sup>*a*</sup> binding site found by MetalPDB server,  $^{50 \ b}$  binding site found by rcsb.org  $^{51}$ 

4AQO for $Ca^{2+}$					
cluster	probability	population	TM-score	ave. ion displacement [Å]	SD
Ι	31.11 %	86	0.55	7.22	3.52
II	24.69 %	62	0.36	9.31	8.42
III	23.98 %	69	0.50	6.52	3.52
IV	14.79 %	35	0.46	11.38	12.22
V	5.42 %	13	0.52	2.99	2.19
6EKG for $Mg^{2+}$					
cluster	probability	population	TM-score	ave. ion displacement [Å]	SD
Ι	27.81 %	56	0.49	26.41	13.43
II	21.98 %	52	0.61	17.93	9.26
III	18.06 %	43	0.57	21.90	5.81
IV	17.39 %	42	0.45	16.02	4.85
V	14.77 %	27	0.45	21.13	6.08
4RJ9 for $K^+$					
cluster	probability	population	TM-score	ave. ion displacement [Å]	SD
Ι	32.95 %	29	0.48	57.11 / 57.39	33.78 / 32.84
II	25.58 %	29	0.48	56.16 / 55.68	38.93 / 39.00
III	19.93 %	26	0.31	65.43 / 65.87	44.30 / 45.32
IV	15.09 %	18	0.51	65.05 / 65.18	38.12 / 36.73
V	6.46 %	5	0.46	5.99 / 10.27	2.52 / 3.28
	I	Į	5E56 for $Na^+$	-	
cluster	probability	population	TM-score	ave. ion displacement [Å]	SD
Ι	29.65 %, 34.15 % <sup><i>a</i></sup>	38, 66 <sup>a</sup>	$0.45, 0.32^{a}$	78.34	23.45
II	22.31 %, 31.08 % <sup><i>a</i></sup>	31, 87 <sup><i>a</i></sup>	$0.43, 0.35^{a}$	79.18	25.45
III	19.60 %, 14.07 % <sup>a</sup>	23, 30 <sup><i>a</i></sup>	$0.48, 0.26^{a}$	76.94	16.77
IV	15.42 %, 12.06 % <sup><i>a</i></sup>	19, 30 <sup><i>a</i></sup>	$0.59, 0.49^{a}$	69.21	18.59
V	13.03 %, 8.64 % <sup>a</sup>	17, 21 <sup><i>a</i></sup>	0.49 <i>,</i> 0.46 <sup><i>a</i></sup>	77.57	18.55
$2Z2K$ for $Cl^-$					
cluster	probability	population	TM-score	ave. ion displacement [Å]	SD
Ι	37.84 %, 23.91 % <sup><i>a</i></sup>	65, 30 <sup><i>a</i></sup>	$0.40, 0.47^{a}$	87.02	22.66
II	23.53 %, 21.32 % <sup><i>a</i></sup>	47, 33 <sup><i>a</i></sup>	$0.48, 0.55^{a}$	80.12	20.40
III	21.05 %, 20.33 % <sup><i>a</i></sup>	28, 44 <sup><i>a</i></sup>	$0.49, 0.57^{a}$	79.85	18.57
IV	16.63 %, 17.41 % <sup><i>a</i></sup>	43, 39 <sup><i>a</i></sup>	$0.46, 0.47^{a}$	81.78	21.45
V	0.95 %, 17.03 % <sup>a</sup>	6, 47 <sup><i>a</i></sup>	0.20, 0.47 <sup><i>a</i></sup>	75.30	16.93

Table 3: Results from simulations started from the experimental structure.

<sup>*a*</sup> simulations in higher ion concentration

Table 4: The experimental apparent equilibrium constants for the binding of metal ions to bovine  $\alpha$ -lactalbumin at 20 C ( $K_{app}$ ) and calculated association and dissociation rate constants ( $k_1$ ,  $k_{-1}$ ) and binding equilibrium constant ( $K_a$ ).

	$k_1$	$k_{-1}$	Ka	$K_{app}[M^{-1}]^{52}$
$Ca^{2+}$	0.6875	0.0038	179.0364	$10^{8.6}$
$Mg^{2+}$	0.7210	0.0042	169.0007	2000
$Na^+$	0.4997	0.4884	1.0231	100
$K^+$	0.4996	0.6184	0.8080	8













## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

• SI.pdf