Predicting non-specifically bound ions

Supported by NIGMS, NIH, grant number: R01 GM093937

Lead scientist: Mrs. Margo Petukh

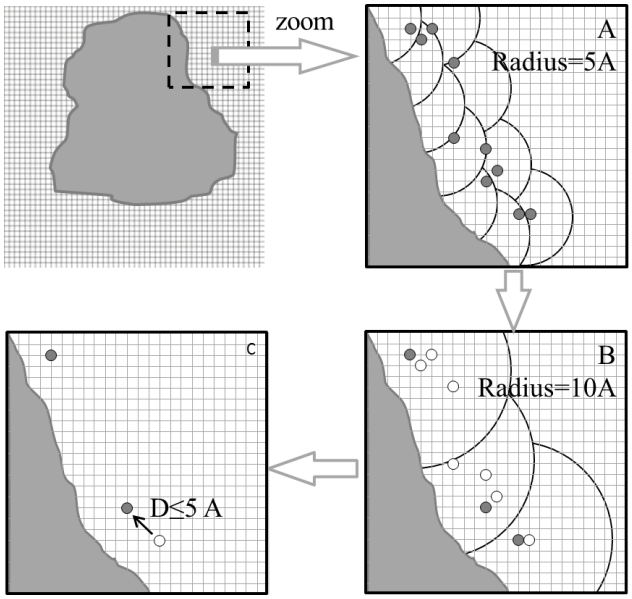
Webserver implementation: Mr. Taylor Kimmett

Abstract: Ions are an important component of the cell and affect the corresponding biological macromolecules either via direct binding or as a screening ion cloud. While some ion binding is highly specific and frequently associated with the function of the macromolecule, other ions bind to the protein surface non-specifically presumably because of electrostatic attraction being strong enough to immobilize them. Here we test such a scenario and demonstrate that experimentally identified surface bound ions are located at potential facilitating the binding, which indicates that the major driving force is electrostatics. Without taking into consideration geometrical factors and structural fluctuations, we show that ions tend to be bound onto the protein surface at positions with strong potential, but with opposite polarity with respect to the ion's polarity. Although such an approach does not distinguish between the chemical nature of the ions, but just their polarity, it can predict non-specific binding of positively or negatively charged ions with acceptable accuracy. These predicted ions can be explicitly treated in the Poisson-Boltzmann (PB) calculations of macromolecules and thus to extend the limits of the PB approach by reducing the magnitude of the local potential.

Method: A typical protein calculated with the above parameters with DelPhi could result in grid size of hundreds and more, which in turn results in more than million grid points were the potential is calculated. Direct analysis of such large array could result in ranking on the top of the list grid points close in space and neglecting other potentially important sites. To avoid such cases, a clustering algorithm was applied.

All points in potential map were analyzed for being on the surface of protein and have negative potential for Mg, Ca and Zn and positive potential for Cl. Points were considered to be on the surface of protein if the shortest distance between them and atoms of the protein is larger than specific vdW bond for tested ion and shorter than 5 Å. Then the grid points were clustered beginning with the point with the smallest X-coordinate and forming a cluster with a radius 5 Å (see figure on the left). Clusters were not intersected. The point with highest absolute potential was chosen to represent the cluster (in figure each representative point for each cluster is shown in black). Once the

representative points for all clusters were determined, we performed sequential clustering to avoid cases of representative points situated in close proximity (Figure B). It was required that the distance between representative points do not exceed 10 Å, and the distance between geometric average of all points in the cluster and furthest to it point in the cluster not to be greater than 5 Å (Figure B, where representative points of clusters are black). If representative points of the clusters with highest absolute potential were situated within 5 Å from each other they were merged to one point with highest absolute potential (Figure 1, C, solid black circles). All points found by cluster method were checked on the ability of appropriate ions to be placed at their position and have SASA from 50% to 75% of the maximum SASA typical for the given type of ion. For that purpose the coordinates of points were artificially added into corresponded protein PDB-file as heteroatoms with ions name.



Results: For each protein in the dataset, the electrostatic potential map was analyzed and the potential clustered as described in the method section. The corresponding representative grid points were ranked by descending absolute value of the potential, so that the point with highest absolute potential has *Rank* =1. Depending on the size, shape, net charge of investigated protein and in general the distribution of the charges inside it, the corresponding electrostatic potential clustering resulted in different number of representative grid points. Figure below shows the distribution of number of representative grid points for each type of ion in proteins from examined dataset (*dark bars*). Significant difference is observed among cases involving Ca and Mg ions (which distribution is broad) versus Zn and Cl ions (which distribution is narrow with a mean about 20-30 representative points). Perhaps, this reflects the difference of the biophysical properties (number of residues, shape and charges) of the corresponding proteins in our dataset which hold different ion type. The same figure (Figure below, *light bars*) shows the *Rank* distribution of the closest (Dmin) to the actual ion's position representative grid point. It should be clarified that due to clustering procedure and the grid algorithm the representative grid points do not necessary have to match the ion's position. It can be seen that in all cases, both for positively and negatively charged ions, the representative grid point closest to the ion's position is ranked among the top 10 points in 30-60 % of the cases (Figure below). The best results were obtained for Mg and Zn atoms, resulting in a sharp peak at distances less than 10 Å. These results indicate that the representative point closest to the actual ions position (Dmin) is

within the top ten representative points with highest electrostatic potential in about 60 % of the cases in the data set. The result for other two types of ions, Ca and Cl, are less impressive, but still indicate a clear trend that the position of the ion binding is within the vicinity of the strongest electrostatic potential.