2

Electrostatics in Protein Engineering and Design

I. John Khan, James A. Stapleton, Douglas Pike and Vikas Nanda
University of Medicine and Dentistry of New Jersey, Piscataway, NJ, USA

1. Introduction

The electrostatic interactions between charged atoms in natural proteins play a central role in specifying protein topology, modulating stability of the molecule, and allowing for the impressive catalytic properties of enzymes. In this chapter, we discuss how protein engineers use the principles of electrostatics and computational protein modeling to develop new proteins for biomedical and biotechnological applications. First, a general introduction is given to familiarize the reader with the important factors to consider in protein electrostatics, and the nature of these electrostatic forces. The next section describes various levels of theory used for modeling electrostatics in proteins. The last sections focus on specific applications in two conceptual classes: the engineering of ionic interactions (1) on protein surfaces, and (2) within the hydrophobic protein core. In both cases, the aim is to promote stability or to control molecular recognition.

2. Important factors influencing protein electrostatics

Interacting ionic species undergo rearrangement of their charge distributions under the influence of each other and their local environment. In electrostatics, we consider the static electrical field that is formed between these charged species once charge rearrangement has occurred. In the context of a protein, this amounts to looking at the many interactions among the polar and/or charged residues scattered throughout the three dimensional structure. Uncharged polar residues can form hydrogen bonding interactions with the hydroxyl (serine and threonine) and amide (asparagine and glutamine) hydrogen bond donors and acceptors on their side chains. Ionizable, or charged, residues have the following titratable side groups: carboxyl (asparate and glutamate), sulfhydryl (cysteine), hydroxyl (tyrosine), guanidino (arginine), amino (lysine), and imidazole (histidine). The ionization state of a titratable residue depends on its pK_a value or proton affinity, which represents the pH at which there is equilibrium between the neutral and charged forms of their respective functional groups.

Electrostatic interactions with the local environment influence the pK_a values of titratable residues. These factors are manifested in the following relationship for the pK_a of a buried residue (Bashford and Karplus 1990; Kaushik *et al.* 2006):

$$pK_a = pK_{a,model} + \Delta pK_{desolv} + \Delta pK_{back} + \Delta pK_{coulomb} \quad (1)$$

The pK_a value for each model residue ($pK_{a,model}$) has been experimentally determined. This value represents the pK_a of the residue when it is completely surrounded by water. Adjustments to the $pK_{a,model}$ as the residue becomes buried are due to the following three factors. The first is the ΔpK_{desolv}, which is the change in pK_a due to the unfavorable removal of an ionized residue from water to the hydrophobic protein core (i.e., a desolvation penalty). The second is the ΔpK_{back}, which is the change in pK_a due to interactions of the buried ionized residue with background charges present within the protein. Background charges are defined as the partial charges of atoms that are manifested as either permanent or induced dipoles in molecules. Examples of background charges are the permanent dipole of a water molecule or the permanent dipole that is formed by an α–helical domain of the protein. The third is the $\Delta pK_{coulomb}$, which is the change in pK_a due to charge-charge interactions among buried ionized residues (or with metal ions, if present). The signs of these ΔpK values will depend on whether the ionized residue is positive or negative, and on the strength of the electrostatic interactions.

The dominant forces to consider in protein electrostatics are the ion-ion, hydrogen bonding, ion-permanent dipole, and permanent dipole-permanent dipole interactions. The strength of these interactions are distance-dependent, as shown in Table 1, with the force of ion-ion pairing being exerted over a significantly longer range compared to weak non-electrostatic forces. For example, the electrostatic force between two charged residues Lys^+ and Glu^- decreases over a distance as $1/r$, whereas the van der Waals attraction between uncharged atoms decreases over a distance as $1/r^6$, where r is on the order of atomic distance. The attraction between the oppositely charged residues, such as Lys^+ and Glu^-, forms a salt bridge, where by definition, both the centroids of their side groups and the charged atoms lie within a range of 4-8 Å (Kumar and Nussinov 2002). Salt bridges, hydrogen bonding, and background charges are commonplace in protein structures, yet proteins are stabilized not only by these electrostatic forces but also by non-electrostatic interactions as well. Examples of non-electrostatic interactions are hydrophobic interactions, van der Waals interactions, disulfide bridges, or covalent bonds.

Type of interaction		**Example**	**Distance dependence**
Ionic	Electrostatic	Lys^+ --- Glu^- (salt bridge)	$1/r$
Hydrogen bonding	Electrostatic	Ser --- carboxyl of peptide bond	Bond length ~ 2.7 Å
Ionic/dipole	Electrostatic	Asp^- --- H_2O	$1/r^2$
Dipole/dipole (permanent)	Electrostatic	Helix dipole --- helix dipole	$1/r^3$
Dipole/dipole (induced)	Non-electrostatic	Dispersion forces	$1/r^6$

Table 1. Relative range of electrostatic and non-electrostatic interactions in proteins

Two other important factors influencing electrostatics in proteins are (1) the dielectric properties of the protein and its surrounding aqueous environment, and (2) the ionic strength of the aqueous environment. The dielectric coefficient (ε) is an indication of polarizability – how readily dipoles can reorient within the medium. In an aqueous

environment, the dipoles of water molecules are free to reorient, hence the dielectric coefficient of water is relatively high with a value of ~78. In contrast, the dielectric coefficient of the protein is lower due to the limited mobility of the protein chain and the nonpolar nature of many amino acid residues. The dielectric coefficient within a protein varies with location, with values of 2-4 for regions having residues that are virtually inaccessible to water (i.e., the hydrophobic core), increasing to values of ~37 near the surface of the protein (Anslyn and Dougherty 2006). As a rule of thumb, we consider the range of electrostatic interactions to be dependent on the dielectric property of the medium according to the plots shown in Figure 1. For example, the energy between point charges in water (ε = 78) cannot be discriminated from baseline thermal energy at a separation of ~2 Å as a result of the charge screening by dipoles of water molecules. In the region below the protein surface (ε = 10) the effective separation increases to ~ 14Å, and within the hydrophobic core (ε = 4) the effective range can be greater than 30 Å. This difference implies that polar and charged residues have greater electrostatic potential when they are buried within the protein. The other factor influencing electrostatic interactions is the ionic strength which also has a screening effect of charge, particularly at the surface of the protein.

Here we focus on the treatment of electrostatics in protein engineering design. For a more general discussion of electrostatics in proteins, we refer the reader to several excellent reviews (Neves-Petersen and Petersen 2003; Bosshard *et al.* 2004; Jelesarov and Karshikoff 2009; Pace *et al.* 2009; Kukic and Nielsen 2010).

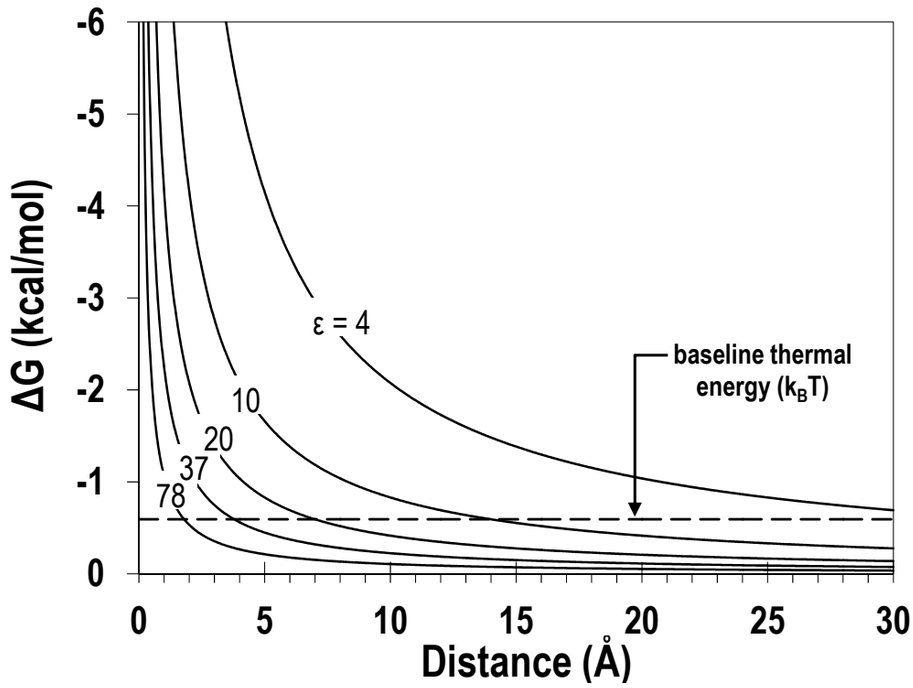


Fig. 1. The change in free energy (ΔG) associated with separating two point charges of opposite sign when surrounded by a medium of dielectric coefficient ε in the absence of salt. The values are calculated using Coulomb's law based on point charges of +0.5 and -0.5. The dashed line represents the baseline thermal energy at 298 K, k_BT, where k_B is the Boltzmann constant and T is the temperature.

3. Theory and modeling of electrostatics in protein engineering

Mechanical modeling of protein structure and dynamics requires a *force field*, a set of atomic and inter-atomic parameters that define how atoms interact in the context of the macromolecule. Among these parameters are the radius and partial charge of specific atom types, optimal bond lengths for atom pairs or optimal angles between groups of three atoms. These parameters are combined in an *objective energy function* which includes bonding (covalent) and nonbonding (electrostatics, van der Waals, hydrogen bonding) interactions to reflect the stability of a specific protein configuration:

$$E = \sum_{bonds} K_r\left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos\left(n\varphi - \gamma\right)\right] + \sum_{atoms\, i<j}\left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}}\right] \quad (2)$$

The first three summations incorporate harmonic or periodic potentials for bond vibrations, bond angle constraints and dihedral (bond rotation) constraints. The final term describes nonbonding interactions, including a van der Waals term that prevents atomic clashes, and an electrostatic term. A detailed explanation of the terms and coefficients for this objective energy function can be found in Cornell (Cornell *et al.* 1995).

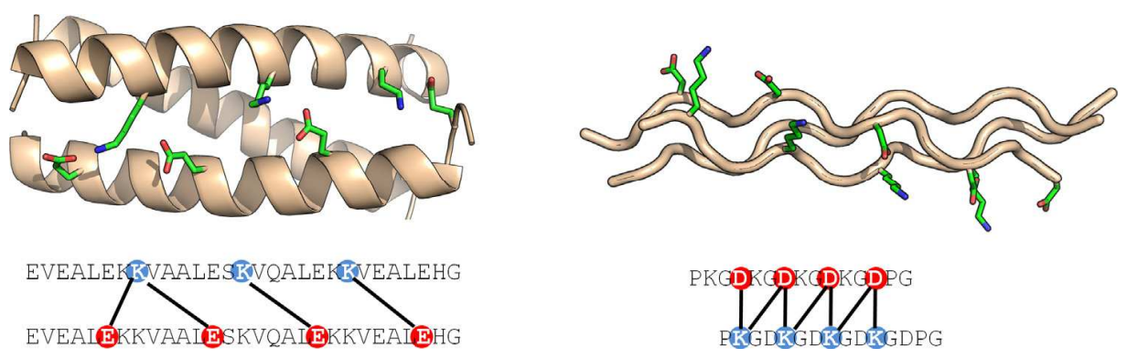


Fig. 2. Charge-pair interactions can be inferred from the sequence for fibrous proteins with periodic structure. (LEFT) The seven-residue heptad of the repeat of the α-helix coiled-coil places acidic (red) and basic (blue) amino acids adjacent in structure. In this case, the interaction between two chains of a three-chain homotrimeric protein are shown (Ogihara *et al.* 1997). (RIGHT) The collagen triple-helix is another type of periodic structure where charge-pairs adjacent in structure can be inferred directly from the sequence. A theoretical model structure of two chains in the triple helix are shown highlighting an extensive charge-pair network. E = glutamic acid, D = aspartic acid, K = lysine.

The objective energy function can be used to predict the lowest energy configuration of a protein chain and to model molecular motions over short (nanosecond to microsecond) periods of time. Electrostatic interactions in the protein are calculated as the sum of all pair-wise atomic interactions. In treating electrostatics this way, two assumptions are made: the partial charge is located at the center of mass of the atom (point charge approximation), and other terms of the multipole expansion beyond ion-ion interactions are disregarded. As

physical chemical methods for measuring protein structure, thermodynamic stability and molecular motions improve, the corresponding force fields that model these processes are updated, improving their predictive power. In protein engineering, the modeling task becomes significantly more difficult relative to molecular dynamics, as we are not only concerned with determining the optimal molecular configuration, but also varying the amino acid sequence to modulate properties of the protein. Even for small proteins, the number of possible sequences to consider is immense: for example, a 100-residue protein has 20 possible amino acid choices at each position, resulting in a total space of $20^{100} \approx 10^{130}$ sequences. Combine this with the configurational degrees of freedom of the protein chain and it is clear that an enumeration of all possible states is computationally impossible. To circumvent this search problem, a number of simplifications or coarse graining approaches are used, and no single level of chemical accuracy is universally applied to all protein engineering problems. The requirements of the design problem dictate the level of theory to use. We present several models of electrostatics with varying levels of chemical accuracy that are employed in protein engineering.

The simplest electrostatics treatments do not incorporate atomic detail and assign discrete values to classes of interactions. This reduces a three-dimensional modeling problem to one dimension and is most useful in the design of molecules where positions that are adjacent in structure can be directly inferred from the amino acid sequence. This scenario is found in fibrous proteins such as α-helical bundles and collagen fibrils (Spek *et al.* 1998). Due to the structural periodicity of the α-helix and the collagen triple-helix, it is possible to anticipate which sequence positions are adjacent in structure (Figure 2). Using this information, a scoring function can be used to optimize these interactions. In many cases, the interactions are designed manually without computation(Berger *et al.* 1996; Lombardi *et al.* 1996; Bryson *et al.* 1998; Olson *et al.* 2001; Shi *et al.* 2001). Amino acids of opposite charge are introduced at adjacent positions such that the maximal number of charge pairs is satisfied. When the design goal is of sufficient complexity that computational intervention is required discrete scores are assigned to interactions (Nautiyal *et al.* 1995). One simple scoring function recently applied to both collagen and α-helical proteins is:

$$\begin{aligned} &\text{Arg / Arg} \quad +2 \\ &\text{Glu / Glu} \quad +3 \\ &\text{Arg / Glu} \quad -1 \end{aligned} \tag{3}$$

In this scenario, any structurally adjacent arginine (Arg) pairs are penalized by two kcals/mole. The penalty for adjacent glutamates (Glu) is greater in anticipation of their shorter side chains which bring repulsive charges in closer proximity. Only favorable Arg/Glu interactions are rewarded. The total energy for a given sequence is the sum of all residue-pair scores. If the number of sequences to sample is small, sequence-space can be fully searched. For larger design problems, Monte Carlo methods such as simulated evolution are often used (Hellinga and Richards 1994). Because they ignore molecular details, these scores are far from accurate, but they allow the rapid evaluation of large ensembles of sequences. The discrete scoring function in equation 3 has been used to design stable helical oligomers with specific composition – e.g. combining α-helical chains A and B yielded an A_2B_2 tetramer, without forming A_4 or B_4 species (Summa *et al.* 2002). The same scoring function was recently extended to design of collagen heterotrimers where three

peptides A, B and C combine specifically to form an ABC heterotrimer (Xu *et al.* 2011). Molecules such as these are now finding applications as synthetic biomaterials where the electrostatic control of self-assembly is responsible for directing the formation of protein fibers (Pandya *et al.* 2000; O'Leary *et al.* 2011).

When it is necessary to include some level of atomic detail in modeling ion-ion interactions, the simplest potential is Coulomb's law:

$$E = 332 \cdot \frac{\bar{q}_i \bar{q}_j}{\varepsilon \cdot r_{ij}} \tag{4}$$

where the interactions of atoms *i* and *j* are a function of charge *q* and the distance of separation *r*. The constant of 332 converts the units of energy to kcals/mole. This can be applied to all atoms in the protein as described in equation 2, or restricted to side chains with a net formal charge. In full-atom implementations, the point charge is located at the center-of-mass of the atom, whereas residue-level charges are often placed at the center of the chemical moiety carrying the partial charge, i.e. the center of the guanidino group of the arginine sidechain. The choice of charge is determined by the force field used.

The strength of a charge-charge interaction is influenced by the polarity of the surrounding medium which is reflected by the choice of dielectric coefficient used. In cases where the structural context is known, often a fixed constant dielectric (e.g. 5-10 for the protein interior and ~78 for the surface) is used. One empirical approximation is to use a distance-dependent dielectric ($\varepsilon = 40 \cdot r_{ij}$) based on the premise that the greater the separation between atoms, the more solvent can access the intervening space and screen electrostatic forces (Mayo *et al.* 1990; Gordon *et al.* 1999). In cases where it is desirable to include the effect of counterions, Debye-Huckel and Coulombic terms can be combined to include an ionic strength parameter (Lee *et al.* 2002).

In addition to charge-charge interactions within the protein, solvent-protein interactions are an important electrostatic component of the free energy of folding. Burial of charged side chains in the protein core comes at the cost of desolvating the sidechain ion. These energies can be modeled with reasonable accuracy using finite difference methods applied to the Poisson-Boltzmann equation (Sharp and Honig 1990), but are infrequently used in protein design applications due to the computational burden. Many software packages dedicated to protein design use an atom or residue-level solvation energy that scales with the fraction of accessible surface area buried upon folding. Although these are grossly approximate calculations, rapid algorithms for calculating solvent-exposed surface area make them attractive for evaluating large numbers of candidate sequences. A number of analytic and empirical methods continue to be developed that are finding applications in protein modeling and design (Flohil *et al.* 2002; Morozov *et al.* 2003; Pokala and Handel 2004; Jaramillo and Wodak 2005; am Busch *et al.* 2008). The assumption that atoms have point charges localized to the atom center of mass becomes problematic when designing proteins where electronic polarizability is important, such as the design of metalloproteins where the solvent reorganization energy around the metal can be important for tuning redox properties (Papoian *et al.* 2003), and enzymes where accurate modeling of the transition state and surrounding ligands is critical for an effective design (Tantillo *et al.* 1998). In this case,

the use of quantum mechanics calculations is warranted. This approach has been used in the engineering of novel protein catalysts where the active site and substrate transition state are modeled using semi-empirical density functional methods (DFT), and the remaining protein treated using standard molecular mechanics and knowledge based potentials (Jiang *et al.* 2008; Rothlisberger *et al.* 2008).

Modeling hydrogen bonding with reasonable accuracy is an important challenge in protein design. Although primarily electrostatic in nature, hydrogen bonds also have partial covalent character which mediates their linearity in molecular structures. Such properties can only be modeled using quantum mechanical (QM) methods, which is computationally infeasible as these are distributed throughout the protein. Instead, empirical functions are often used that include both proximity and orientation terms. These have been refined using the extensive database of high-resolution protein structures to develop knowledge-based potentials that can capture subtle properties (Kortemme *et al.* 2003). QM methods can also be used to explore the role of other types of electrostatic interactions such as cation-π interactions between ions and aromatic amino acids. In the next sections, several examples of protein engineering of electrostatic properties are presented, highlighting the application of various levels of theory as needed to achieve the design objective.

4. Surface charges in protein electrostatics

It was long thought that surface electrostatics do not make a significant contribution to protein stability because the interactions of polar residues with water in the unfolded state are as energetically favorable as their interactions with each other in the folded state. However, recent work has demonstrated that surface charge optimization can offer significant stability increases to a wide range of proteins (Schweiker and Makhatadze 2009). Surface charge optimization is an attractive option for protein engineering and design because surface positions are generally much more permissive to mutation compared to buried positions, where side chains are prone to clashing as they pack tightly into the protein core. Nature also takes advantage of this evolutionary flexibility at surface positions, modifying surface charge interactions to modulate energetic folding barriers (Halskau *et al.* 2008) and to stabilize thermophilic proteins.

The hypothesis that surface electrostatics can be important for stability is supported by the observation that thermophilic proteins generally contain more charged surface residues than their mesophilic analogs (Kumar and Nussinov 2001). Thermophilic proteins have evolved to be active at high temperatures, and their structures must therefore be very stable. This stabilization is achieved through a number of different strategies, including enriching the sequences in charged surface residues and buried hydrophobic residues at the expense of polar residues. This adaptive response to evolutionary pressure for increased stability has been reproduced in computer simulations of simple lattice model proteins (Berezovsky *et al.* 2007). As a result, the number of salt bridges in a protein is correlated with the temperature of the environment in which its host organism lives (Kumar *et al.* 2000). In one study, mutations to two surface residues of a mesophilic cold shock protein (one of which eliminated an unfavorable electrostatic interaction) yielded a mutant that nearly matched the stability of the thermophilic version of the protein (Perl *et al.* 2000). The stability change was greatly reduced in the presence of 2M NaCl, confirming the importance of electrostatic interactions (which are sensitive to the screening effects of salt) in stabilizing the mutant.

Designed and engineered proteins can also benefit from the stability gains that are possible by optimizing surface electrostatics. The generality of this strategy for protein stabilization was demonstrated experimentally in a study in which the surface residues of a diverse set of five proteins were modified (Strickler *et al.* 2006). A computational algorithm was used to search for mutations to surface positions that would provide the maximum improvement to the energy by adding favorable interactions or alleviating unfavorable ones. Because the combinatorial space of possible surface charges is too large to cover exhaustively, a genetic algorithm was used to search for near-optimal sequences. Genetic algorithms efficiently sample sequence space by mimicking the natural evolutionary process. A population of sequences is generated and evaluated with an energy function - in this case, the energy function was based on a solvent accessibility-corrected Tanford-Kirkwood model. The top-scoring sequences are kept, multiplied, and diversified by random mutations within sequences and crossover or recombination events in which sections are swapped among multiple sequences. At the end of the process, sequences containing between three and eight mutations were selected. One to three designs were constructed for each target protein, synthesized, and purified. Protein unfolding was then monitored by circular dichroism spectroscopy. Remarkably, an increase in stability relative to the wild-type was observed for each of the designed sequences. The largest increase in stability was 4.4 kcal/mol. Another recent study applied this approach to the surface electrostatics optimization of two enzymes. The activity of enzymes is often highly sensitive to even small perturbations to the active site. Nonetheless, human acylphosphatase (AcPh) and human cell-division cycle 42 factor (Cdc42) were successfully stabilized by surface charge optimization with no loss in enzymatic activity (Gribenko *et al.* 2009). Mutant sequences were chosen that maximized the improvement in electrostatic energy while limiting the number of mutations from the wild-type sequence to ~5% of the total residues. The stability of each modified protein was ~10°C higher than their corresponding wild-type protein, while the structures, monomeric nature, and enzymatic activities were retained. This study demonstrated the possibility of increasing the stability of an enzyme by making rational mutations to surface residues on the basis of electrostatic calculations, without disturbing the protein core or the enzymatic activity.

In addition to influencing the intramolecular stability of engineered proteins, electrostatics are important in intermolecular interactions. The balance of charged and hydrophobic residues in a protein sequence is important in determining the tendency of that sequence to aggregate when unfolded (Calamai *et al.* 2003; Chiti *et al.* 2003; Pawar *et al.* 2005). Charged residues within otherwise hydrophobic regions can act as "sequence breakers" that prevent those regions from aggregating. The ability of like-charge repulsive interactions to discourage aggregation is the basis of a surface electrostatics engineering strategy called "supercharging" (Lawrence *et al.* 2007). Amino acids at surface positions of a supercharged protein are mutated to charged residues so that the net charge of the protein is maximized. Net positive and net negative supercharged proteins have both been shown to be less prone to aggregation than their corresponding wild-types. For example, the green fluorescent protein (GFP) is unable to refold into a fluorescent state after thermal denaturation because of aggregation with neighboring unfolded chains. However, the extremely high net charges of supercharged GFP chains disfavor interactions with other unfolded chains of like charge (Figure 3). When a GFP variant supercharged to a net charge of +36 was thermally or chemically denatured, the sample was able to regain up to 62% of its initial fluorescence, confirming that the high net charge of the protein disfavored interchain aggregation.

Streptavidin and glutathione-S-transferase were also successfully supercharged to yield highly aggregation-resistant engineered variants. The supercharging process often involves a relatively large number of mutations, but because the hydrophobic core of the protein is undisturbed, protein folding is typically not significantly adversely affected. For example, the -7 net charge of superfolder GFP was pushed to extremes of +48 (by 36 mutations) or -30 (by 15 mutations). Remarkably, despite the repulsion that would be expected from gathering so many like charges on the surface of a protein, and the stability to be gained by optimizing surface charges demonstrated by the studies presented earlier, the supercharged GFPs were able to fold and fluoresce normally, with only slight decreases in thermodynamic stability. The destabilizing effect of the high concentration of like charges at the surface may be limited by equal or greater destabilization of competing states within the denatured state ensemble (Pace *et al.* 2000). The intuitive electrostatics-based supercharging strategy has already become a popular choice among protein engineers for stabilizing *de novo* designed proteins and therapeutic peptides against aggregation.

Another major limitation of peptide therapeutics is the difficulty of transporting peptides and proteins across the cell membrane. Currently, the leading strategy to improve cellular uptake is to express the target protein as a fusion with one of several polycationic amino acid sequences derived from natural cell-penetrating peptides (Heitz *et al.* 2009). In a recent study, positively supercharged GFP was shown to be capable of entering a range of mammalian cells, and of delivering fused protein payloads more effectively than the standard cationic fusion tags Tat, Arg_{10}, and penetratin (Cronican *et al.* 2010).

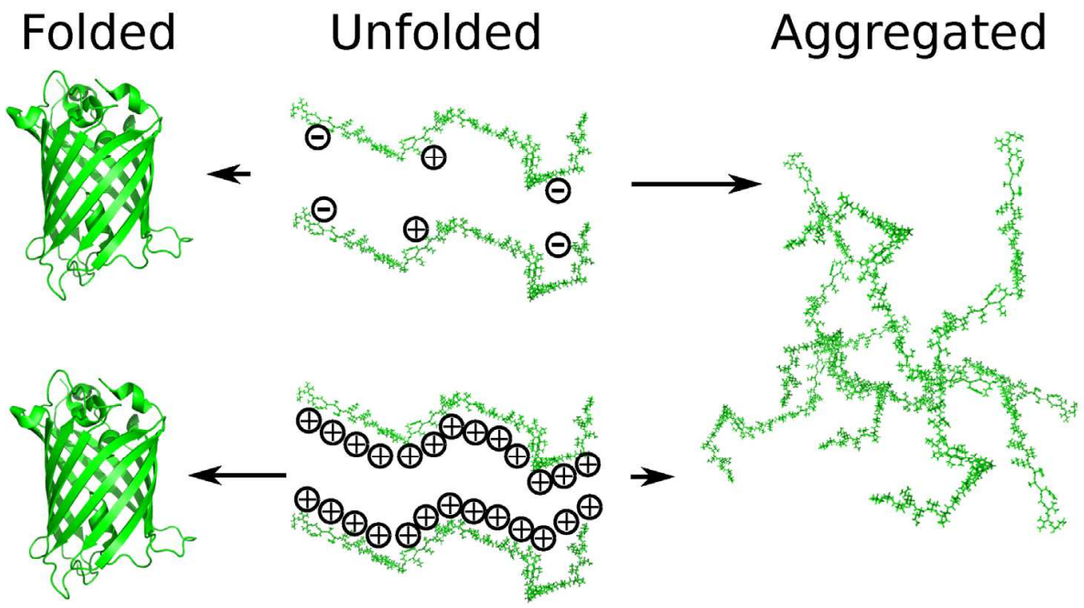


Fig. 3. Supercharging decreases the tendency of unfolded proteins to aggregate by increasing like-charge repulsion. Thermally denatured green fluorescent protein (center) is capable of refolding into the fluorescent state (left) or aggregating with other unfolded chains (right). In the case of the wild-type protein (top), aggregation dominates. In contrast, a sample of a supercharged version of GFP with a net charge of +36 (bottom) regained 62% of its fluorescence following thermal denaturation. The like-charge repulsion between the positive charges on each denatured supercharged polypeptide mitigated aggregation.

5. Electrostatics with buried polar or charged residues

Proteins can tolerate the burial of ionizable residues when environmental modification of the pK_a of the buried side chains prevents them from assuming the charged state. In a series of studies using 96 variants of an engineered form of staphylococcal nuclease, hydrophobic buried residues were individually mutated to lysine, glutamate, or aspartate (Isom *et al.* 2008; Isom *et al.* 2010; Isom *et al.* 2011). The apparent pK_a values of these residues were determined by curve fitting plots of the changes in free energy associated with individually charging the mutants relative to a reference state as a function of pH. In general, the pK_a values of these buried residues were shifted by the environment so that they existed in neutral form within the hydrophobic core.

Protein function can be improved by the burial of a polar residue if the conformation of an associated ligand can be stabilized by electrostatic interactions. Enhanced cyan fluorescent protein (ECFP) was optimized as a FRET[1] donor molecule by mutating several of its residues – S72A[2], Y145A and H148D (Rizzo *et al.* 2004; Malo *et al.* 2007). The new protein variant was called Cerulean. The authors describe the contribution of the H148D substitution of Cerulean in stabilizing a single conformation (i.e., the *cis*-form) of its associated chromophore. Unlike the histidine residue in ECFP, the buried aspartate side group stabilized the *cis*-conformation of the internal chromophore as part of an extended network of hydrogen bonding which included forming a bifurcated hydrogen bond with the indole nitrogen of the chromophore (Figure 4A). The pK_a of the buried aspartate was estimated to be ~6 allowing the residue to remain protonated (i.e., neutral form) for hydrogen bonding, and the smaller size of the aspartate (relative to the histidine) aided in packing of the core. Other hydrogen-bonding interactions were made with nearby polar side groups and with bound water which provided a cage-like enclosure for the internal chromophore (not shown in the figure). The *cis*-conformation of the chromophore placed the six-membered ring of the indole in close proximity to the imidazolinone ring, enhancing energy transfer. The result of the H148D substitution was an engineered molecule that had relatively homogeneous exponential fluorescence emission decay, a property which is necessary for fluorescence-lifetime measurement studies.

The burial of an ionized residue in a protein is an unfavorable event that can be countered by stabilizing electrostatic interactions such as the formation of hydrogen bonding networks. The enzyme ribonuclease T1 is an example of a protein that contains an ionized buried residue, D76, that lacks an ion-pairing partner with which it can form a stabilizing salt bridge. The measured pK_a of D76 is extremely low (pK_a~0.5), ensuring that it always remains fully charged. As a result, it forms a hydrogen bonding network with nearby polar residues T91, Y11, and N9, and with bound water molecules in the protein (Giletto and Pace 1999). This local conformation is depicted in Figure 4B. The wild-type ribonuclease has been shown to have better thermal and chemical stability when compared to uncharged variants D76N, D76S and D76A of the enzyme. In this instance, having a buried charge within a polar microenvironment is advantageous.

[1]FRET = Förster Resonance Energy Transfer
[2]The standard one-letter code is used to designate the amino acid residues

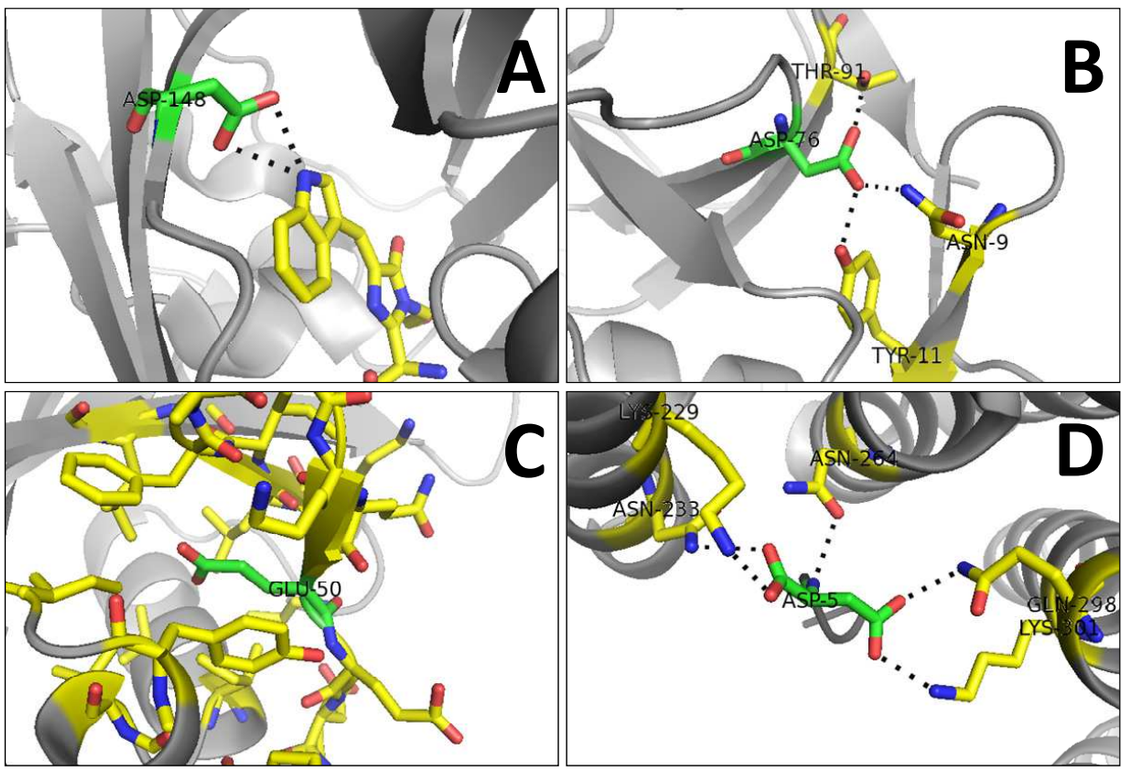


Fig. 4. (A) The key placement of a buried titratable residue can enhance protein function. The cis-conformation of the indole ring of the Cerulean chromophore (shown in yellow) is stabilized by a substituted buried aspartate (shown in green) (PDB code 2q57). The chromophore is comprised of two rings, an indole and an imidazolinone, connected by a methylene bridge. The structure is further stabilized by a network of hydrogen bonding with backbone residues and bound water molecules surrounding the chromophore (not shown). (B) A buried ionized residue that is unable to form a salt-bridge can be stabilized by a hydrogen-bonding network. A charged aspartate residue (shown in green) is stabilized by a network of hydrogen bonding among three polar residues (shown in yellow) within the hydrophobic core of ribonuclease T1 (PDB code 9rnt). (C) The burial of a charged residue can be used to destabilize the protein structure. Glutamate is substituted for leucine (L50E; shown in green) within the hydrophobic core of ubiquitin (PDB code 1ubq). The buried glutamate is surrounded by a hydrophobic microenvironment (shown as yellow residues within 8 Å). The ionization of glutamate results in unfavorable conditions for the charged residue leading to local unfolding in the protein. This charge burial strategy was used to stabilize high-energy folding intermediates of ubiquitin. (D) Residues that become buried following protein-protein interaction can form stabilizing hydrogen-bonding networks. A buried two-carboxylate aspartate of Hsp90 C-terminal peptide (shown in green) is stabilized through its interactions with the polar residues of HOP (shown in yellow) at the protein interface (PDB code 1elr). The Hsp90 peptide is further stabilized along its length by hydrogen bonding with the side chains of the HOP helices (not shown). All figures are generated with PyMOL (Schrodinger, LLC) using a color scheme of red for oxygen and blue for nitrogen, and black dotted lines are used to indicate hydrogen bonding. Hydrogen atoms are not explicitly shown.

Native proteins can be unfolded by ionized residues buried within the hydrophobic core if there are no stabilizing electrostatic interactions to counter the charge. This notion was exploited in a charge burial strategy where "foldons" (regions of secondary structures that cooperatively unfold) of ubiquitin were selectively destabilized in order to trap high-energy intermediate folded states of the protein (Zheng and Sosnick 2010). A strategically located hydrophobic buried residue was substituted with glutamate which was subsequently ionized (Glu^-) during pH titration. In the case of an L50E substitution located at the C-terminal end of the β5 strand, the Glu^- was placed in a microenvironment that was dominantly hydrophobic, with no nearby polar residues or backbone nitrogens to stabilize the charge (Figure 4C). Structural change within the protein caused by the Glu^- was monitored by nuclear magnetic resonance spectroscopy where the authors were able to detect the sequential unfolding of the β5 strand and an adjacent 3_{10}-helix. Interestingly, these unfolded intermediates could be stabilized by pH, and it was possible to refold the protein back to its native structure by neutralizing Glu^-.

Protein-protein interfaces rely on electrostatic interactions to stabilize their previously exposed charged or polar residues. A study on the binding interaction between heat shock protein (Hsp)-organizing protein (HOP) domain TPR2A and the C-terminal end of Hsp90 (MEEVD) revealed the formation of an extensive network of hydrogen bonding between the ionized residues on Hsp90 and polar groups on TPR2A (Kajander *et al.* 2009). As an example, we illustrate the stabilization of the two carboxyl groups of the C-terminal aspartate residue, which is clamped by polar side chains from the TPR2A α-helices and forms hydrogen bonds with K229, N233, Q298 and K301 (Figure 4D). Of these polar groups, N233 was found to be one of several significant binding surface residues that become buried. Similar electrostatic interactions were found along the length of the binding cavity, demonstrating how interfacial residues are stabilized.

6. Acknowledgements

This work was supported by grants from the National Institute of Health (5R21AI088627, 5R01GM089949, 1DP2OD006478 and 1F32GM099291) and the National Science Foundation (DMR0907273).

7. References

am Busch, M. S., A. Lopes, et al. (2008). Testing the Coulomb/Accessible Surface Area solvent model for protein stability, ligand binding, and protein design. BMC Bioinformatics 9, 148.

Anslyn, E. V. and D. A. Dougherty (2006). Modern physical organic chemistry. Sausalito, CA, University Science.

Bashford, D. and M. Karplus (1990). Pkas of Ionizable Groups in Proteins - Atomic Detail from a Continuum Electrostatic Model. Biochemistry 29(44), 10219-10225.

Berezovsky, I. N., K. B. Zeldovich, et al. (2007). Positive and negative design in stability and thermal adaptation of natural proteins. PLoS Comput Biol 3(3), e52.

Berger, J. S., J. A. Ernst, et al. (1996). Stabilization of helical peptides by mixed spaced salt bridges. Journal of Biomolecular Structure & Dynamics 14(3), 285-291.

Bosshard, H. R., D. N. Marti, et al. (2004). Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. J Mol Recognit 17(1), 1-16.

Bryson, J. W., J. R. Desjarlais, et al. (1998). From coiled coils to small globular proteins: design of a native-like three-helix bundle. Protein Sci 7(6), 1404-1414.

Calamai, M., N. Taddei, et al. (2003). Relative influence of hydrophobicity and net charge in the aggregation of two homologous proteins. Biochemistry 42(51), 15078-15083.

Chiti, F., M. Stefani, et al. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 424(6950), 805-808.

Cornell, W. D., P. Cieplak, et al. (1995). A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. J Am Chem Soc 117(19), 5179-5197.

Cronican, J. J., D. B. Thompson, et al. (2010). Potent delivery of functional proteins into Mammalian cells in vitro and in vivo using a supercharged protein. ACS Chem Biol 5(8), 747-752.

Flohil, J. A., G. Vriend, et al. (2002). Completion and refinement of 3-D homology models with restricted molecular dynamics: Application to targets 47, 58, and 111 in the CASP modeling competition and posterior analysis. Proteins-Structure Function and Genetics 48(4), 593-604.

Giletto, A. and C. N. Pace (1999). Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T-1. Biochemistry 38(40), 13379-13384.

Gordon, D. B., S. A. Marshall, et al. (1999). Energy functions for protein design. Current Opinion in Structural Biology 9(4), 509-513.

Gribenko, A. V., M. M. Patel, et al. (2009). Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. Proc Natl Acad Sci U S A 106(8), 2601-2606.

Halskau, O., Jr., R. Perez-Jimenez, et al. (2008). Large-scale modulation of thermodynamic protein folding barriers linked to electrostatics. Proc Natl Acad Sci U S A 105(25), 8625-8630.

Heitz, F., M. C. Morris, et al. (2009). Twenty years of cell-penetrating peptides: from molecular mechanisms to therapeutics. Br J Pharmacol 157(2), 195-206.

Hellinga, H. W. and F. M. Richards (1994). Optimal Sequence Selection in Proteins of Known Structure by Simulated Evolution. Proceedings of the National Academy of Sciences of the United States of America 91(13), 5803-5807.

Isom, D. G., B. R. Cannon, et al. (2008). High tolerance for ionizable residues in the hydrophobic interior of proteins. Proceedings of the National Academy of Sciences of the United States of America 105(46), 17784-17788.

Isom, D. G., C. A. Castaneda, et al. (2011). Large shifts in pK(a) values of lysine residues buried inside a protein. Proceedings of the National Academy of Sciences of the United States of America 108(13), 5260-5265.

Isom, D. G., C. A. Castaneda, et al. (2010). Charges in the hydrophobic interior of proteins. Proceedings of the National Academy of Sciences of the United States of America 107(37), 16096-16100.

Jaramillo, A. and S. J. Wodak (2005). Computational protein design is a challenge for implicit solvation models. Biophys J 88(1), 156-171.

Jelesarov, I. and A. Karshikoff (2009). Defining the role of salt bridges in protein stability. Methods Mol Biol 490, 227-260.

Jiang, L., E. A. Althoff, et al. (2008). De novo computational design of retro-aldol enzymes. Science 319(5868), 1387-1391.

Kajander, T., J. N. Sachs, et al. (2009). Electrostatic Interactions of Hsp-organizing Protein Tetratricopeptide Domains with Hsp70 and Hsp90 COMPUTATIONAL ANALYSIS AND PROTEIN ENGINEERING. Journal of Biological Chemistry 284(37), 25364-25374.

Kaushik, J. K., S. Iimura, et al. (2006). Completely buried, non-ion-paired glutamic acid contributes favorably to the conformational stability of pyrrolidone carboxyl peptidases from hyperthermophiles. Biochemistry 45(23), 7100-7112.

Kortemme, T., A. V. Morozov, et al. (2003). An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. Journal of Molecular Biology 326(4), 1239-1259.

Kukic, P. and J. E. Nielsen (2010). Electrostatics in proteins and protein-ligand complexes. Future Med Chem 2(4), 647-666.

Kumar, S. and R. Nussinov (2001). How do thermophilic proteins deal with heat? Cell Mol Life Sci 58(9), 1216-1233.

Kumar, S. and R. Nussinov (2002). Close-range electrostatic interactions in proteins. Chembiochem 3(7), 604-617.

Kumar, S., C. J. Tsai, et al. (2000). Factors enhancing protein thermostability. Protein Eng 13(3), 179-191.

Lawrence, M. S., K. J. Phillips, et al. (2007). Supercharging proteins can impart unusual resilience. J Am Chem Soc 129(33), 10110-10112.

Lee, K. K., C. A. Fitch, et al. (2002). Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein. Protein Sci 11(5), 1004-1016.

Lombardi, A., J. W. Bryson, et al. (1996). De novo design of heterotrimeric coiled coils. Biopolymers 40(5), 495-504.

Malo, G. D., L. J. Pouwels, et al. (2007). X-ray structure of cerulean GFP: A tryptophan-based chromophore useful for fluorescence lifetime imaging. Biochemistry 46(35), 9865-9873.

Mayo, S. L., B. D. Olafson, et al. (1990). Dreiding - a Generic Force-Field for Molecular Simulations. Journal of Physical Chemistry 94(26), 8897-8909.

Morozov, A. V., T. Kortemme, et al. (2003). Evaluation of models of electrostatic interactions in proteins. Journal of Physical Chemistry B 107(9), 2075-2090.

Nautiyal, S., D. N. Woolfson, et al. (1995). A designed heterotrimeric coiled coil. Biochemistry 34(37), 11645-11651.

Neves-Petersen, M. T. and S. B. Petersen (2003). Protein electrostatics: a review of the equations and methods used to model electrostatic equations in biomolecules--applications in biotechnology. Biotechnol Annu Rev 9, 315-395.

O'Leary, L. E., J. A. Fallas, et al. (2011). Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. Nat Chem 3(10), 821-828.

Ogihara, N. L., M. S. Weiss, et al. (1997). The crystal structure of the designed trimeric coiled coil coil-VaLd: implications for engineering crystals and supramolecular assemblies. Protein Sci 6(1), 80-88.

Olson, C. A., E. J. Spek, et al. (2001). Cooperative helix stabilization by complex Arg-Glu salt bridges. Proteins-Structure Function and Genetics 44(2), 123-132.

Pace, C. N., R. W. Alston, et al. (2000). Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. Protein Sci 9(7), 1395-1398.

Pace, C. N., G. R. Grimsley, et al. (2009). Protein ionizable groups: pK values and their contribution to protein stability and solubility. Journal of Biological Chemistry 284(20), 13285-13289.

Pandya, M. J., G. M. Spooner, et al. (2000). Sticky-end assembly of a designed peptide fiber provides insight into protein fibrillogenesis. Biochemistry 39(30), 8728-8734.

Papoian, G. A., W. F. DeGrado, et al. (2003). Probing the configurational space of a metalloprotein core: an ab initio molecular dynamics study of Duo Ferro 1 binuclear Zn cofactor. J Am Chem Soc 125(2), 560-569.

Pawar, A. P., K. F. Dubay, et al. (2005). Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. Journal of Molecular Biology 350(2), 379-392.

Perl, D., U. Mueller, et al. (2000). Two exposed amino acid residues confer thermostability on a cold shock protein. Nat Struct Biol 7(5), 380-383.

Pokala, N. and T. M. Handel (2004). Energy functions for protein design I: efficient and accurate continuum electrostatics and solvation. Protein Sci 13(4), 925-936.

Rizzo, M. A., G. H. Springer, et al. (2004). An improved cyan fluorescent protein variant useful for FRET. Nature Biotechnology 22(4), 445-449.

Rothlisberger, D., O. Khersonsky, et al. (2008). Novel Kemp Elimination Catalysts by Computational Enzyme Design. Nature.

Schweiker, K. L. and G. I. Makhatadze (2009). A computational approach for the rational design of stable proteins and enzymes: optimization of surface charge-charge interactions. Methods Enzymol 454, 175-211.

Sharp, K. A. and B. Honig (1990). Calculating Total Electrostatic Energies with the Nonlinear Poisson-Boltzmann Equation. J. Phys. Chem. 94, 7684-7692.

Shi, Z., C. A. Olson, et al. (2001). Stabilization of alpha-helix structure by polar side-chain interactions: complex salt bridges, cation-pi interactions, and C-H em leader O H-bonds. Biopolymers 60(5), 366-380.

Spek, E. J., A. H. Bui, et al. (1998). Surface salt bridges stabilize the GCN4 leucine zipper. Protein Science 7(11), 2431-2437.

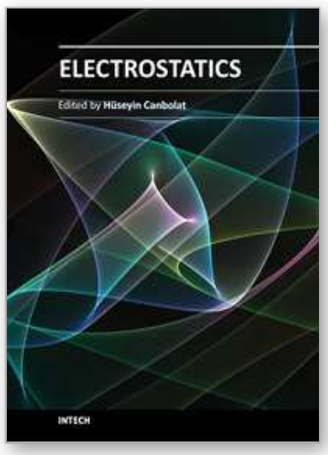
Strickler, S. S., A. V. Gribenko, et al. (2006). Protein stability and surface electrostatics: a charged relationship. Biochemistry 45(9), 2761-2766.

Summa, C. M., M. M. Rosenblatt, et al. (2002). Computational de novo design, and characterization of an A(2)B(2) diiron protein. J Mol Biol 321(5), 923-938.

Tantillo, D. J., J. Chen, et al. (1998). Theozymes and compuzymes: theoretical models for biological catalysis. Current Opinion in Chemical Biology 2, 743-750.

Xu, F., S. Zahid, et al. (2011). Computational design of a collagen a:B:C-type heterotrimer. J Am Chem Soc 133(39), 15260-15263.

Zheng, Z. and T. R. Sosnick (2010). Protein vivisection reveals elusive intermediates in folding. Journal of Molecular Biology 397(3), 777-788.



Electrostatics
Edited by Dr. Hüseyin Canbolat

ISBN 978-953-51-0239-7
Hard cover, 150 pages
Publisher InTech
Published online 14, March, 2012
Published in print edition March, 2012

In this book, the authors provide state-of-the-art research studies on electrostatic principles or include the electrostatic phenomena as an important factor. The chapters cover diverse subjects, such as biotechnology, bioengineering, actuation of MEMS, measurement and nanoelectronics. Hopefully, the interested readers will benefit from the book in their studies. It is probable that the presented studies will lead the researchers to develop new ideas to conduct their research.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

I. John Khan, James A. Stapleton, Douglas Pike and Vikas Nanda (2012). Electrostatics in Protein Engineering and Design, Electrostatics, Dr. Hüseyin Canbolat (Ed.), ISBN: 978-953-51-0239-7, InTech, Available from: http://www.intechopen.com/books/electrostatics/electrostatics-in-protein-engineering-and-design

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen