Review

Protein electrostatics: From computational and structural analysis to discovery of functional fingerprints and biotechnological design

Filippo Vascona,a, Matteo Gasparottoa, Marta Giacomellob,c, Laura Cendrona, Elisabetta Bergantinoa, Francesco Filippinia,⇑ Irene Righettoa,⇑

a Synthetic Biology and Biotechnology Unit, Department of Biology, University of Padua, Italy
b Bioenergetic Organelles Unit, Department of Biology, University of Padua, Italy
c Department of Biomedical Sciences, University of Padua, Italy

Article info
Article history:
Received 16 March 2020
Received in revised form 16 June 2020
Accepted 19 June 2020
Available online 30 June 2020

Abstract
Computationally driven engineering of proteins aims to allow them to withstand an extended range of conditions and to mediate modified or novel functions. Therefore, it is crucial to the biotechnological industry, to biomedicine and to afford new challenges in environmental sciences, such as biocatalysis for green chemistry and bioremediation. In order to achieve these goals, it is important to clarify molecular mechanisms underlying proteins stability and modulating their interactions. So far, much attention has been given to hydrophobic and polar packing interactions and stability of the protein core. In contrast, the role of electrostatics and, in particular, of surface interactions has received less attention. However, electrostatics plays a pivotal role along the whole life cycle of a protein, since early folding steps to maturation, and it is involved in the regulation of protein localization and interactions with other cellular or artificial molecules. Short- and long-range electrostatic interactions, together with other forces, provide essential guidance cues in molecular and macromolecular assembly. We report here on methods for computing protein electrostatics and for individual or comparative analysis able to sort proteins by electrostatic similarity. Then, we provide examples of electrostatic analysis and fingerprints in natural protein evolution and in biotechnological design, in fields as diverse as biocatalysis, antibody and nanobody engineering, drug design and delivery, molecular virology, nanotechnology and regenerative medicine.

© 2020 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

Prologue: scope and contents of this mini-review ......................................................... 1775
1. Electrostatics: a pivotal player in protein structure and interactions .......................... 1775
1.1. Protein folding and stability .................................................................................... 1775
1.2. Enzyme catalysis .................................................................................................. 1777
1.3. Protein-protein interactions .................................................................................. 1777
1.4. Protein binding to other molecules ....................................................................... 1778
1.5. Electrostatic interactions in protein binding to plasma membrane and organelles ............................ 1778
2. Computational analysis and prediction tools ............................................................. 1779
2.1. Basis of continuum electrostatic calculations ....................................................... 1779
2.2. Prediction tools .................................................................................................... 1780
2.2.1. APBS ................................................................................................................. 1780
2.2.2. WebPIPSA ....................................................................................................... 1780
2.2.4. PBEQ-Solver ................................................................................................. 1781

⇑ Corresponding author.
E-mail addresses: francesco.filippini@unipd.it (F. Filippini), irene.righetto@bio.unipd.it (I. Righetto).
1 Equal contribution.

https://doi.org/10.1016/j.csbj.2020.06.029
2001-0370/© 2020 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Prologue: scope and contents of this mini-review

Several excellent reviews and exhaustive chapters can be found in literature about electrostatics theoretical background and its influence on protein properties [1–5], therefore, the need for one more may seem futile. However, the aim of such articles was “to present a comprehensive coverage on the uses of electrostatic interactions by proteins in tuning their basic biophysical properties and in performing their biological functions” [1]. However, according to the scope of this journal, we aim to highlight how structural computational studies focusing on protein electrostatics may boost applied science and biotechnology projects. In this mini-review, we provide examples of bioinformatic tools used in prediction-driven projects that led, via the computational analysis of electrostatics, to the successful engineering of proteins or the identification of functional fingerprints. Indeed, apart from its crucial role in mediating basic functions, protein electrostatics can also be studied for biotechnological purposes, i.e. to provide information on its role in modulating enzyme or antibody/nanobody features of industrial interest (e.g. solubility, catalytic efficacy, binding strength and specificity), protein-nanoparticle interactions exploited in regenerative medicine and drug delivery, mechanisms underlying subcellular sorting to membranes and organelles and viral antigenic drift and host jump, which are highly important to human and animal health and vaccine design.

Section 1 and its subsections depict the overall role of electrostatics in modulating protein stability and function; computational methods and tools that could be used to infer functional insights are reviewed in section 2, while section 3 focuses on aforementioned examples of applied science/biotechnology projects in which protein electrostatics investigation played a central role.

1. Electrostatics: a pivotal player in protein structure and interactions

Each protein consists of a peculiar combination of amino acids, which, in turn, can be classified according to shared features, e.g hydrophobic/polar, small/large, presence/absence of a positive/negative charge. Hydrophobic and electrostatic interactions influence the fate of any protein since its synthesis and folding to its degradation. Hereafter, we shortly summarize the main processes in which protein electrostatics play a role. In subsections 1.1 and 1.2, intra-chain electrostatics and interactions between protein residues and the solvent are considered as determinants for proper folding and stability, as well as a pivotal player in enzyme catalysis; then, we report about the role of electrostatics in regulating protein interactions with other proteins (subsection 1.3) or with molecules other than proteins, either cellular/natural or not (subsection 1.4). An overall picture of protein electrostatics is presented in Fig. 1.

1.1. Protein folding and stability

Proper protein folding and stability are maintained by many interactions between inner and outer residues. Traditionally, hydrophobic effect, hydrogen bonding, and packing interactions between buried protein residues have been considered as the dominant ones, while the role of surface features – particularly electrostatic ones – has often been neglected [3]. Although electrostatic interactions are long in range, and thus weakly specific, they can contribute to proper folding and stability e.g. via salt bridges or by further modulating the folding landscape.

The relevance of electrostatics in the interactions of proteins with other polypeptides or non-proteinaceous molecules is widely acknowledged (see next subsections). In contrast, the role it can play in establishing specifically funnelled landscapes and their relevance in determining the final protein structure (and its conformational changes) is still unclear [2]. The influence of electrostatic forces on proteins arises as early as during translational events. The ribosomal exit tunnel is negatively charged, but the local potential through its length could be influenced by charged amino acids of nascent peptides as they move along the tunnel, generating a wave of electrostatic potential; this, in turn, regulates secondary structure acquisition and protein elongation rates and pauses, drastically contributing to nascent peptide folding or misfolding [6]. Furthermore, the ribosome outer surface is also characterized by a diffused negative electrostatic potential (mainly due to rRNA), and there is some evidence pointing out at its interaction with charged residues of nascent polypeptides, in turn affecting their conformational sampling ability, cotranslational events including interactions, aggregation and degradation [7] and ultimately folding kinetics [8].

In addition to synthesis, folding, and maturation, electrostatics is involved in protein degradation. Although the majority of proteins is degraded through a ubiquitin dependent mechanism, several direct proteasome signals (DPSs) exist that can mediate proteasome-directed protein degradation in a ubiquitin independent manner.

Recently, Kudriavtseva and co-workers [9] characterized the myelin basic protein (MBP) DPS, showing that it is enriched in basic and flexible amino acids and its signaling function is charge-mediated. Notably, MBP mutants with negative or less positive surface charge showed a decrease in proteasomal degradation rate in vitro. It seems possible that MBP high cationic charge may function as a built-in ubiquitin-mimetic signal that can be attracted by the proteasome. The resulting interaction could be mediated by ionic complexes between basic amino acids in MBP and anionic residues of the proteasome.

The role played by electrostatic forces in governing non-specific protein-protein interactions and stability has been investigated by “supercharging” mutation analysis [10]. Even though such an approach can be of interest to biotechnology projects (see section...
Fig. 1. Examples of interactions mediated by electrostatic forces. (a) Intramolecular salt bridges are involved in protein folding and stability; (b) Cotranslational interaction with ribosomes may influence secondary structure formation; (c) Long-range interaction between charged residues is involved in tertiary structure acquisition; (d) Surface charge can act as a direct proteasome signal and trigger protein degradation; (e) Electrostatic can facilitate complex formation; (f) Conformational exchange; (g) Positive charge is required for rycin binding to rRNA (h) Protein binding to DNA could hamper binding of other molecules modulating DNA accessibility.
The relationship between net charge and refolding ability still needs complete elucidation, as diverse sets of supercharging mutations have different outcomes, likely because of the specific structural contributions given by each residue [11].

The impact of electrostatics on protein stability is very complex because of the competition between several phenomena, including local dipolar interactions, attraction and repulsion between charged residues and desolvation penalties of their ionized groups (which depend on the solvent dielectric properties, in turn affected by the temperature [4,12]). Given the ionizable nature of such residues, pH also influences protein stability [13]. Furthermore, the balance between electrostatic and hydrophobic contributions should be taken into account when dealing with protein stability. As a result of this complexity, many studies seem to point out conflicting results on the stabilization/destabilization potency of electrostatic interactions. For example, the aforementioned supercharging strategy for improving protein refolding goes along with the observation of an increased amount of charged residues and a high number of salt bridges and hydrogen bonds in thermostable proteins [12]. On the other hand, this is in contrast to the abundance of highly charged proteins in aggregates found in aging organisms and to their proposed pronounced susceptibility to destabilization by age-related oxidation and charge modification [14]. Furthermore, some works calculated a destabilizing effect of ionizable groups due to charged residues desolvation during protein folding [5] and to like charge repulsion, especially in proteins with a high net charge [14]. In this context, amino acids position and surrounding environment are critical: as a rule of thumb to avoid a destabilizing effect. Hydrophobic residues are usually buried into the protein core whereas polar and charged amino acids are more often solvent-exposed [1], unless charge burial into protein core entails the establishment of favorable interactions (not necessarily ion-pairing [13]). Analogously, nonpolar and charged residues are usually positioned in the core and at the rim of interacting protein interfaces, respectively [1]. Interestingly, Kumar and Nussinov [12] proposed that the notable amount of buried salt bridges found in hyperthermophilic proteins maintain a stabilizing effect since many cooperative electrostatic interactions are established both between the salt-bridging side chains and between these residues and others inside the protein, without suffering from solvent screening.

### 1.2. Enzyme catalysis

Understanding how enzymes work remains one of the challenges of modern biophysics [15]. Even though electrostatic effects are the best candidates to explain enzyme catalysis, early studies in this field did not consider this possibility, likely because studies on reactions in solution focused on rate constants rather than on rate-limiting activation barriers. In the absence of a proper computational model, it is almost impossible to assess the dielectric effects in the protein and thus to estimate the strength of electrostatic effects. However, pioneering work [16] provided the first quantitative hint that electrostatic effects may have a major role in enzyme catalysis, suggesting that the catalytic power of enzymes is almost exclusively due to electrostatic effects and specifically to the preorganized electrostatic environment of their active sites [15,17]. Since all quantum chemistry is electrostatic, “electrostatic catalysis” [16] takes into account the effects of the protein charges, permanent/induced dipoles, metal ions and the solvation by bound water molecules [15]. Studies that identified electrostatic effects as a key factor in enzyme catalysis concerned several enzymes and enzyme classes, as reviewed by Warshel and co-workers [15]. In addition to the direct effect on catalysis mediated by transition-state stabilization, electrostatics also influences the allosteric control of catalytic efficiency. For instance, in the activation of Ras, GAP binding results in a major electrostatic stabilization favoring GTP hydrolysis by an arginine finger and by stabilization of the catalytic configuration [15]. The importance for catalysis of electrostatic preorganization over conformational motions could be depicted by accurate computational analyses. In particular, a work on dihydrofolate reductase and other enzymes demonstrated that the role of flexibility and conformational dynamics in catalysis is negligible, while the largest contribution arises from electrostatic preorganization, as changes in motion depend in turn on changes in the reaction potential surface, modifying the reorganization free energy [18,19].

Krzeminska and co-workers investigated the dynamic and electrostatic effects governing the catalytic power of HIV-1 PR protease, by computing the free energy surfaces for all possible mechanisms, in terms of potentials of mean force. Such calculations showed that the contribution of dynamic effects to the activation free energy is relatively low, while the electric field created in the active site is critical for the electronic reorganization required during the reaction [20].

Computational combined quantum mechanics and molecular mechanics (QM/MM) methods are currently considered as the most appropriate to simulate dynamics and to dissect steps of chemical catalysis as it proceeds in very large and flexible systems such as enzymes. These methods have been applied to the reaction catalyzed by a rat-liver glycine N-methyltransferase (GNMT), that is the transfer of a methyl group from the S atom of the donor molecule S-adenosyl-L-methionine (SAM) to the acceptor N atom of a glycine. The wild-type enzyme and three mutants, in which a single active-site tyrosine had been changed either in alanine, glycine or phenylalanine, were compared to check whether mechanical compression by the protein structure could destabilize the reactant state more than the transition state (resulting in a lower free-energy barrier). Slight differences were observed, in accordance with similar calculated values of activation free energy, but mutants showed a diminished ability to achieve a determined donor-acceptor distance. Mechanical constrains were consequently ruled out and results were interpreted by suggesting the WT enzyme provided the most favorable electrostatic environment to stabilize the charge of the methyl group all along its transfer (i.e., as the reaction proceeds, in the stabilization of the transition state with respect to the reactant state, in the accommodation of the donor-acceptor distance, in each conformational fluctuation till the release of the product) [21].

### 1.3. Protein-protein interactions

Electrostatic forces play an established role in protein-protein interactions (PPIs). The electrostatic interaction energy between two molecules carrying only one unit of net charge and positioned 10 Å away from each other is much higher (at such distances) than any other energy component contributing to binding [22]. Of course, in addition to electrostatic forces depending on residues of protein partners, salt concentration and pH also play an important role in modulating electrostatics guidance of PPIs, as reviewed by Jensen [23]. The intracellular environment is populated by a huge crowd of molecules, including proteins, each needing to find the right partner among hundreds of thousands of candidates from the cellular proteomic complement; this notwithstanding, a fast recognition process is needed in most instances [24]. Indeed, the recognition process can be enhanced by long-range electrostatic guidance, a force that selects and brings the interacting partners together. A few examples, among many others, of PPIs favored by electrostatic steering, are interactions between (i) fasciculin and acetylclophilinesterase [25], (ii) formation of α-lactalbumin–l-lysine heterodimers [26], (iii) antibody Fc heterodimer formation in bispecific antibodies [27], and Cdc42 recognition by Wiskott-
Aldrich syndrome proteins [28]. Formation of the protein complex KIX-pKID is favored by electrostatic interactions providing charge-charge stabilization of the native state and contributing to the funneling of the binding landscape [2].

Chaperones recognition of unfolded proteins is commonly believed to rely mostly on hydrophobic interactions [29]; without diminishing the importance of hydrophobic effects, long-range electrostatic interactions have been recently found to play a crucial role in the first binding stage of some molecular chaperones (e.g. Escherichia coli Spy) to their client proteins [29]. In contrast, other chaperones (i.e. Dnak of E. coli) not only benefit from electrostatic steering effects but specifically recognize patterns including charged residues amongst nonpolar ones [30].

Methods for driving alteration of the electrostatic properties are considered for designing proteins with optimized binding and activity [31]; for instance, electrostatic optimization was used for modulating protein-protein association rates [32]. As reported by Ritchie and Webb [33], structural and electrostatic factors are crucial in the affinity and specificity of macromolecular interactions and chemical reactivity. For example, specific electrostatic interactions between charged amino acid residues regulate the binding of (i) von Willebrand factor to blood platelets [34] and (ii) of GTP binding proteins of the RAS superfamily with downstream effector proteins [33]. It is well-known that clusters of charged and polar residues that are located at protein-protein interfaces may enhance complex stability [35].

1.4. Protein binding to other molecules

Electrostatics plays a central role in the interaction of proteins with highly charged biological molecules such as DNA, RNA and membrane phospholipids. For instance, it is widely known that most nucleic-acid binding proteins are rich in positively charged residues (especially in their binding regions), which mediate attraction with the negatively charged backbone of DNA/RNA. Intriguingly, several nucleic acid-binding proteins may cause frustration in the DNA-binding region, which favors their binding to DNA but hampers binding of other protein partners. These interactions are mainly driven by electrostatic forces that, therefore, contribute to modulating nucleic acid accessibility [2].

Long- and short-range interactions play a pivotal role in the assembly of the spliceosomal U1 small nuclear ribonucleoprotein. In particular, long-range electrostatic interactions mediate high-affinity binding between the positive charge of U1A protein and the negatively charged stem loop 2 RNA of the U1 snRNA, while short-range interactions (hydrogen bonds among RNA bases and amino acid side chains) favor a specific binding [36].

DCL-1 is an intrinsically disordered protein involved in miRNA biogenesis that recognizes the dsRNA by acquiring a well-folded structure after engagement with its interaction partner; electrostatics was found to play a key role in both the protein-RNA recognition and in folding of the complex [37].

3dRPC, an algorithm for prediction of 3D RNA-protein complex structures, has been used for sampling possible RNA-protein complex conformations by calculating the geometric and electrostatic complementarities and stacking interactions at the RNA-protein interface [38].

The long-range character of electrostatic forces has been exploited by molecular recognition processes that need fast rates. As an example, very rapid degradation of some molecules may represent an adaptive advantage, e.g. high rate in neurotransmitters clearance is crucial to fast resetting of neuromuscular junctions and to mediate quick escape from predators and dangerous agents [39]. Indeed, electrostatic steering enhances by few hundred fold the degradation of the positively charged acetylcholine molecule by the negatively charged active-site region of synaptic acetylcholinesterase [40].

Another example of electrostatics involvement in fast biological processes is represented by ricin poisoning. The catalytic A chain of ricin acts as a glycosidase able to remove a specific adenine residue from an exposed loop of the 28S rRNA, leading to rRNA breakage, ribosome inactivation, and final block of protein synthesis. The high toxicity of ricin depends on the capacity of its A chain to inactivate a few thousand ribosomes per minute, i.e. faster than the cell can make new ones (https://www.uniprot.org/uniprot/P02879). The rate constant for this reaction at low ionic strength is increased by roughly a 3 million factor because of electrostatic attraction between ricin and the ribosome [41].

1.5. Electrostatic interactions in protein binding to plasma membrane and organelles

The role of electrostatic interactions as mediator of protein localization at different cell compartments has been studied by several research groups. It has been proposed that peripheral proteins attach to biological membranes either due to the structural and geometric arrangement of the latter, that is curvature and lipid packing, or to electrostatic forces exerted by negatively charged lipids [42]. Coordination of these two “protein adsorption strategies” fundamentally impacts on protein localization at well-defined microdomains and consequently, on the initiation of processes that require biological membranes, like autophagosomes formation and vesicle budding or fusion [42,43]. The three main classes of lipids present in eukaryotic membranes are glycerophospholipids (GPLs), sphingolipids and sterols. The GPL class is the most enriched one, accounting for phosphatidylcholine (PtdCho), phosphatidylethanolamine (PtdEtn), phosphatidylserine (PtdSer) and phosphatidylinositol (PtdIns). The most prominent lipid involved in membrane electrostatics is the negatively charged PtdSer. PtdSer is produced through phospholipid exchange between the Endoplasmic Reticulum and mitochondria at sites where the two organelles lie in close proximity [44,45]. The membranes that define these points of juxtaposition are known as mitochondria associated membranes (MAMs): despite being produced there, PtdSer is not a main component of MAMs, being retrieved at high percentages (up to 10%) on the Plasma Membrane (PM). Along with the phosphoinositide phosphatidylinositol(4,5) biphosphate (PIP2), PtdSer mediates the association of proteins at membranes of the late secretory pathway, which is directly proportional to their electrostatic potential (the extent of their charge) without any dependence on the exact aminoacidic sequence [46]. An interesting example of electrostatically-driven subcellular localization concerns the tumor suppressor PTEN, whose attachment to PM is mediated by polycationic stretches within its sequence. Phosphorylation at its C-terminus hides this positive charge, thus controlling its localization and function [47]. Another example regards the G proteins Rac1/2, which associates to the highly charged surfaces of endosomes and cytosolic side of the PM proportionally to the number of basic residues included at their C-termini [46,48,49]. Bacterial toxins also exploit electrostatics to strive small G proteins, thus exerting their cytotoxic effects [50].

Phagocytic membranes facing the cytosol are enriched of PtdSer and PIP2, playing a role in pathogen engulfment [51]. The negative charge however decreases along the endocytic pathway, due to the gradual decrease of PIP2 abundance from the initial phagosome to lysosomes [48,49,52]. The importance of electrostatics in the phagocytic process is corroborated by two findings. First, some bacteria (e.g. Salmonella typhimurium) have developed as a pro-survival mechanism the expression of a phosphatase able to reduce PIP2 and PtdSer levels, hence impairing the endocytic path and hampering their lysosomal-mediated degradation [53]. Second,
Electrostatic dynamics associated with phagocytosis underlie a recent drug delivery approach based on the use of Cell Penetrating Peptides (CPPs). Some CPPs are enriched of positively charged amino acids which favour the phagocytic internalization of their cargo. While drug internalization with this approach is effective, the success of CPPs-fused therapeutics is blunted because drugs are degraded during the final steps of the phagocytic process. To avoid this issue, secondary/tertiary amine groups having a pKa close the pH of the endosome have been introduced. Their proton buffering ability reduces endosomal vesicle acidification, in turn blocking the extrusion of chloride through endosomal pumps and enhancing water uptake into the vesicle, thus leading to osmotic swelling and drugs release into the cytosol [54,55].

Protein linkage to anionic lipids is pivotal not only for protein localization and related intracellular processes, but also for regulatory feedback loops on the membrane composition itself. Negative surface potential indeed attracts cationic sphingolipid hydrolases/-transferases and phospholipases that, through their activity, lower or increase the abundance of their substrates thus contributing to membrane remodelling [56,57]. Electrostatic interactions at the surface of organelles and cytosolic side of PM are also regulated by scramblases, floppases and flippases, enzymes responsible for the flip-flop of GPLs and hence asymmetry of biological membranes. Reversing the distribution of GPLs between the two sides of the lipid bilayer is yet another signal transduction mechanism. Exposure of the negatively charged PtdSer moiety to the external side of the PM results in cell signaling cascades both intracellularly and at extracellular level and cell death. PtdSer depletion of the PM cytosolic leaflet would induce the release of electrostatically-associated proteins into the cytosol, while its exposure on the surrounding environment would represent a cue for phagocytes recruitment and clearance of dead cell [58,59].

Changes in electrostatic interactions during apoptosis occur not only at PM: negative charges on the surface of mitochondria contribute to the recruitment of cationic proteins, such as K-Ras, Bid and caspase 8. A diatomic molecule, cardiolipin, is exposed on the surface of these organelles during early steps of the apoptotic process: this seems pivotal for the progression of the cell death cascade as dampening of the negative membrane charge through cardiolipin binding proteins blocks the process [60,61].

On the contrary, electrostatics play minor roles in the association of proteins to lipid rafts, which are membrane portions enriched of cholesterol, sphingomyelins, phospholipids with saturated fatty acid chains and gangliosides at the expense of phospholipids with unsaturated acyl chains [62].

Basic electrostatic features (and organellar association) of proteins are further modulated by post-translational modifications, like phosphorylation and lipidation (introducing negative charges), proteolysis that removes charged aminoacidic sequences [63–65].

Overall this evidence highlight that electrostatic interactions are key mediators of protein localization and that keeping the balance between charged and non-charged membranes domains is critical for cell physiology.

### 2. Computational analysis and prediction tools

Solvent plays a pivotal role in mediating biochemical processes such as the interaction between proteins or between proteins and ligands. The environment in which these events take place, is often made up of water and ions. While modeling proteins in such a complex setting, water molecules and ions can be treated in explicit or implicit (continuum) way. The former one allows for more accurate simulation but it is more expensive from a computational point of view. Conversely, implicit solvent methods are affected by drawbacks related to lower accuracy, but are faster and more suitable for studies focusing on the solute (the protein) behavior. Therefore, we choose to discuss only the continuum solvent treatment. Several methods are available in the field and this section is not a comprehensive list; hereafter, instead, we provide an overview of popular methods used in the protein electrostatic computations field, which are user-friendly and suitable for biotechnologists (see Table 1).

| Name             | URL                          | Features                                      | References |
|------------------|------------------------------|-----------------------------------------------|------------|
| APBS             | http://nbcr-222.ucsd.edu/pdb2pqr_2.0.0/ | Electrostatic focusing, pKa calculations       | [72,77,78] |
| WebPIPSA         | http://pipsa.eml.org         | Electrostatic potential comparison             | [73,79]    |
| DelPhi Suite     | http://compbio.clemson.edu/ | Object-oriented, Gaussian-based smooth dielectric function, pKα calculations, DelphiForce | [75]       |
| PBEQ-Solver      | http://www.charmm-gui.org/input/pbeqsolver | Electrostatic focusing, pKa calculations, protein/DNA/RNA electrostatic interaction energy calculation | [76]       |
| Blues            | http://protein.bio.unipd.it/blues/ | Born radii calculations, solvation electrostatic free energy calculations, electrostatic forces calculation, pKa calculations, electrostatic potential calculation | [71]       |

2.1. Basis of continuum electrostatic calculations

Electrostatic properties depend on the distribution of whole and partial charge on the 3D protein structure [31]. Coulomb’s law expresses the electrostatic potential $V(r)$ as follows:

$$V(r) = \frac{1}{4\pi\varepsilon_0} \frac{q}{r}$$

where: $q$ = charge; $r$ = distance; $\varepsilon_0$ = dielectric constant relative to vacuum permittivity and $\varepsilon_0$= dielectric medium. However, Coulomb’s law is not suitable for describing electrostatics in proteins, as it describes a system with a single dielectric medium, whereas proteins show a hydrophobic core enveloped by the solvent. Therefore, electrostatic calculations for proteins are carried out using the Poisson-Boltzmann Equation (PBE):

$$\nabla^2 \psi = \frac{\varepsilon_0}{\varepsilon_r} \frac{\partial \psi}{\partial r}$$

where $\psi$ = electrostatic potential.

Here, the solvent is treated as implicit: in this way, dynamic effects of water are not directly internalized, leading to a better analysis of electrostatics [31]. Protein dielectric constants express the effect of the protein environment, reflecting proteins structure and sequence properties [66]. Moreover, continuum electrostatic studies and molecular dynamic simulations revealed that different structural motifs located in the protein may present different dielectric constant values [67]. Implicit solvent calculations (continuum electrostatics) provide a water phase atomic detail reduction and intrinsically an equilibrium solution [68]. The medium dielectric constant is 80 ($\varepsilon = 80$, water), whereas a dielectric constant of 4 ($\varepsilon = 4$), used for protein, should account for electronic polarization and small backbone fluctuations. Amin and Küpper, in a recent paper calculated the mean dielectric constant of more than 150000 proteins in the PDB, presenting a value of 3.23. Once calculations have been performed (e.g. by software tools reported
Isopotential contours are plotted at levels of different $\frac{z n k_B T}{e}$ which means that the electrostatic component of the potential energy of interaction between the protein field and an elementary charge of $+1e$ located somewhere on the $+n k_B T/e$ isopotential surface, would be equal to:

$$
\left(\frac{n \cdot k_B T}{+1 \cdot 1.602E - 19\text{Coulombs}}\right)
$$

$$
= n \cdot \left(\frac{8.314\text{Joules/Kelvin/6.022E + 23}}{298K}\right)/\left(+1 \cdot 1.602E - 19\text{Coulombs}\right)
$$

$$
= n\text{Volts}
$$

Another formalism involved in continuum electrostatics is Generalized Born (GB) method. This approach also takes part in protein and nucleic acids Molecular Dynamics (MD) simulations in order to model hydration effects and obtain solvent-dependent forces [70]. GB model is an approximation for the linear PBE:

$$
G_s = -\frac{1}{8\pi\varepsilon_0} \left(1 - \frac{1}{\varepsilon_0 - \varepsilon_1}\right) \sum_{y} \frac{q_y}{f_{GB}}
$$

Where

$$
\frac{1}{f_{GB}} = \sqrt{\left(\frac{r_y}{a_y} + a_y^2 e^{(-D)}\right)}
$$

$$
D = \frac{r_y}{2a_y}
$$

$$
a_y = \sqrt{a_i a_j}
$$

where $\varepsilon_0$ is the permittivity of free space, $\varepsilon$ is the dielectric constant of the solvent, $q_y$ is the charge on particle $y$, $r_y$ is the distance between the particles 1 and $y$ and $a_y$ is the Born radius, expressing the distance between the atom and the molecular surface. The Born equation describes the transfer free energy of a single spherical ion having a single charge at its center from the gas phase to an environment characterized by $\varepsilon$. The main goal of GB model relies on the estimation of polarization charge contributions to the self-energy of each charge and the interaction energy of each pair of charges [71].

2.2. Prediction tools

2.2.1. APBS

Adaptive Poisson-Boltzmann Solver (APBS) software package can solve PBE using electrostatic “focusing”. This is a popular finite difference technique for generating accurate solutions to the PBE in subsets of the problem domain, such as a binding or titratable sites within a protein. This approach implies that charges and dielectric constants are discretized over a grid. Protein molecular surface (MS) is mapped onto a user-defined density 3D grid, then used to obtain the finite difference solutions of the PBE. The product of the electrical potential and charge at each voxel (grid point), where a real charge has been mapped, provides the electrostatic free energies. APBS performs calculations by using initially a coarser grid and then a finer one for the refinement. APBS carries out calculations with a grid spacing of 0.5 Å. APBS first defines the solvent accessible regions of the protein and calculates the electrostatic potential for each of the grid points. Finally, the electrostatic potential is mapped onto the MS (.dx file). Electrostatic calculations can be applied on molecules of a wide size range. APBS can be used through PDB2PQR server [72] at http://nbcr-222.ucsd.edu/pdb2pqr_2.0.0

2.2.2. WebPiPSA

WebPiPSA [73] (http://pipsa.eml.org) allows computing and comparing electrostatic potential among a large number of proteins. This pipeline launches a workflow combining several algorithms thus it is not based on a single calculation method. Specifically, after structures upload, WebPiPSA workflow can be summarized as follows:

- **Structures superimposition**: with “sup2pdb” option selected, the sequence of one structure, referred to as the template, undergoes a pairwise sequence alignment with the remaining coordinate files. Alignments are then used to perform structures superimpositions;

  - **Polar hydrogens addition**: WHAT IF [74] adds polar hydrogen atoms to the structures. Protonation is executed at pH 7 for all residues except for His, which is treated as singly or doubly protonated;

  - **Electrostatic potentials calculations**: this step can be performed using APBS or UHBD. Electrostatic potentials are automatically calculated. The user must choose ionic strength and temperature. The solvent is treated as implicit;

  - **Electrostatic potentials comparison**: a probe of radius 2Å defines the protein surface. PiPSA compares potentials in the complete protein surface skins. The skin extends out from the protein surface with a thickness of 3Å. Electrostatic protein comparison is possible due to the implementation of the Hodgkin or Carbo similarity indexes. The similarity indexes range from -1 (anti-correlated potentials) through 0 (uncorrelated) to +1 (identical potentials). These values are converted into distances given by $\sqrt{2 - 2SI}$ where SI stands for similarity index. Distance values are comprised between 0 (identical) and 2 (anti-correlated potentials);

    - **Clustering analysis and epograms generation**: in this last step WebPiPSA generates an output consisting of a heat map (representing the distance matrix) and an epogram, allowing for the fast identification of inter-protein relations.

2.2.3 DelPhi Suite

DelPhi Suite [75] is available at http://compbio.clemson.edu/sapp/delphi_webserver/ or as a standalone software. This popular platform has been redesigned, utilizing the object-oriented programming technique and other unique features provided by C++ to ensure various levels of multiprocessing and memory distribution. This leads to a significant improvement in computational time.
needed to solve the linear and nonlinear PBEs and large-scale modeling. In the case of relatively small systems (less than 300³ grid points), the best choice is the regular single-CPU implementation. For medium-size problems (300³ to 600³ grid points) the OpenMP implementation is the most suitable choice. Finally, extremely large problems (>600³ grid points) can be handled by the MPI DelPhi implementation. Both standalone and webserver versions are characterized by the following features:

- **Gaussian-based smooth dielectric function**: This approach takes into account the inhomogeneity of both the macromolecule and the space between the solute and the solvent. Here, there is a smooth transition between macromolecule and water phase: densities are represented as Gaussian density functions and the dielectric constant as a function of space. This method assigns higher dielectric constant at the protein surface with respect to the hydrophobic core; moreover, the cavities inside the protein are described by a dielectric constant higher than in the rest of the macromolecule but smaller than for bulk water. The Gaussian-based smooth dielectric function shows that surface water molecules are not much freer to move than the bulk ones.

- **Mobile ions treatment**: The treatment of mobile ions relies on a desolvation penalty term in the PBE in order to avoid the ions to enter the protein.

- **Zeta-potential**: It reflects the aggregation tendency. High values of Zeta-potential mean electrostatic repulsion and therefore, no particles/macromolecules aggregation.

- **pKa computational approach**: It is located in a Gaussian-based approach protocol. Here macromolecules ionizable groups pKas are calculated without defining molecular surface (surface-free pka approach). This method is able to predict the pKas of polar groups and it takes into account salt effects.

- **DelPhiForce**: The electrostatic force between two molecules (proteins, DNAs, lipids, small molecules, etc.) is computed via the FRC module. Electrostatic forces are pivotal in both molecules docking and mutual orientation prior to this physical event. The electrostatic force between two objects can be highlighted by vector representations.

- **SAAFEC method**: SAAFEC stands for “Single Amino Acid Folding Free Energy Changes”. DelPhi Suite is able to model the effect of nonsynonymous variants on the folding free energy, particularly important for personalized medicine applications due to the correlation between the magnitude of folding/free energy change and the propensity of a given mutation to be pathogenic. SAAFEC method is based on Molecular Mechanics-Poisson-Boltzmann (MM/PBSA) approach.

- **SAAMBE method**: SAAMBE (Single Amino Acid Mutation related change of Binding Energy) method computes binding free energy changes due to amino acid substitutions. The method is characterized by a MM/PBSA-based component and a set of statistical terms.

- **SAMPDI method**: It computes the effects of amino acids substitutions on the binding free energy of protein-DNA/RNA complexes. This is a MM/PBSA-based feature.

- **MSSM method**: The MultiScale Sampling Method is able to model the docking between large and small molecules via a Monte Carlo procedure. First, MSSM calculates the electrostatic energy of the entire system at a coarse-grained resolution. Then, this information is shifted to a region of interest, calculating the electrostatic energy at a significantly finer resolution.

### 2.2.4. PBEQ-Solver

PBEQ-Solver software [76] (http://www.charmm-gui.org/input/pbeqsolver) carries out PBE calculations using coarse-grid spacing (1.5 Å before and 1 Å after focusing). This program is available as a web GUI, implementing the CHARMM PBEQ module to solve the PB equation. The electrostatic potential of the solvent accessible surface and the isopotential contours are visualized by a Java applet (MarvinSpace molecular visualization software). PBEQ-Solver is able to compute protein-protein/DNA/RNA electrostatic interaction energy and pKa of selected titratable residues in both aqueous and membrane environments. Calculations are provided in two steps:

1. Conversion of a PDB file into CHARMM readable files: the user can select part of a protein chain or a model in the case of NMR structures, modify engineered residues, select terminal group, protonation, disulfide bonds, phosphorylation state, or generate a biologically functional unit or a crystal packing. In this first step, PDB Reader is able to detect and display disulfide bonds and they can be removed. Moreover, the software can spot engineered residues, converting them to the corresponding natural ones. Finally, PDB Reader is able to manage undetermined coordinates by building them using a predetermined internal coordinate table.

2. PB calculations: the PBEQ-Solver module can perform three types of PB calculations for:

   - **Electrostatic potential and salvation free energy**:
     \[ \Delta G_{\text{elec}} = \frac{1}{2} \sum_i q_i \phi_{\text{eff}}(r_i) \]
     
     where \( \Delta G_{\text{elec}} \) is the electrostatic solvation energy and \( \phi_{\text{eff}}(r) \) is the reaction field potential.

   - **Protein-protein/DNA/RNA interaction energy**:
     \[ \Delta G_{\text{inter}} = \frac{1}{2} \left[ \sum_{i \neq j} q_i \phi_{\text{eff}}(r_{ij}) - \sum_{i} q_i \phi_{\text{eff}}^A(r_i) - \sum_{j} q_j \phi_{\text{eff}}^B(r_j) \right] \]
     
     where \( \phi_{\text{eff}}^A(r) \), \( \phi_{\text{eff}}^B(r) \) and \( \phi_{\text{eff}}^C(r) \) are the electrostatic potentials respectively of the complex and its components A and B in solvent or membrane environments.

   - **pKa of a selected titratable residue**:
     \[ pK_{\text{a,shift}} = \frac{\Delta G}{2.303 R T} \]
     
     where \( \Delta G = \Delta G_{\text{prot}} - \Delta G_{\text{model}} \)

### 2.2.5. Bluues

Bluues software [71], available at http://protein.bio.unipd.it/bluues/, performs electrostatic calculations using the GB model. At first, the program computes generalized Born radii from numerical surface integrals. Then, Born radii are used to carry out electrostatic analyses. Depending on user’s requirement, the output will show:

- the generalized Born radius of each atom;
- the solvation electrostatic free energy;
- the electrostatic forces on each atom;
- the pH-dependent properties;
- the pKa of all ionizable groups;
- the electrostatic potential at the surface of the molecule;
- the electrostatic potential in a volume surrounding the molecule.

### 3. From functional fingerprints to biotech applications

Biological activity requires proper interaction between proteins and their interactors. Since electrostatics plays a pivotal role in defining molecular recognition, the assessment of electrostatic complementarity of protein-ligand complexes provides insights on why ligands and on how binding can be modulated. These issues have been addressed by Bauer and Mackey [80], who developed a system to predict electrostatic complementarity (EC) between enzymes and substrates. Particularly, the authors suggest that displaying EC onto molecule surfaces might be a straightforward-
ward method to determine which mutations may improve binding efficiency and, therefore, opening the road for next steps in protein engineering.

Hereafter we provide several examples of biotechnological applications where electrostatics modulation acts as the main player.

3.1. Electrostatic fingerprints in enzyme stability and solubility

Covalent disulfide bridges, representing the strongest intra-chain bonds in proteins, are often crucial to protein thermal stability and thus of special interest in the stabilization of biotechnologically relevant mesophilic enzymes by rational engineering. However, covalent bridges could be too rigid to allow proper conformational changes required for protein function and stabilization can be obtained by modifying the number and positioning of salt bridges, as electrostatic forces are strong enough to boost stabilization, while still allowing for local conformational plasticity needed to mediate biochemical functions.

Stability and solubility are essential characteristics for enzymes used as biocatalysts. In the presently imperative research of new tools and methods for green chemistry applications, attention has mainly been focused on (i) accumulating genomic and metagenomic data [81]; (ii) the engineering of enzymes by rational and/or directed evolution [82]; and, only more recently, on (iii) the development of new compatible ionic liquids (ILs) to be employed in substitution of harmful organic solvents. Advanced ILs are sought in biocatalysis as potential ‘greener’ solvents, being non-volatile, non-flammable, biodegradable, less expensive and generally well-tolerated by proteins [83,84]. In any case, engineering approaches aiming to increase enzyme stability and solubility also have to safeguard (or possibly improve) their enzymatic activity. So far, when computational analysis has been employed as a preliminary or synergic tool to genetic engineering, it has been preferentially focused on electronic interactions occurring in internal shells rather than on the surface of protein structures. Indeed, charges in or nearby the catalytic site (coordinating e.g. metal ions, cofactors, substrate molecules, reaction intermediates), defining its entrance or, also, circumscribing tunnels for access to or exit from it, are the most obvious determinants of activity and stability. In spite of this, charged amino acid distribution on the surface of an enzyme has been almost exclusively considered as a mere determinant of protein solubility. However, surface charges have a key role in mediating the conformational response of enzymes to the chemical-physical behavior of the medium in which the reaction occurs. This response, even though not always evident, in turn, has an unavoidable impact on the catalytic activity, in terms of both kinetic parameters and regio- or stereo-selectivity. A recent example of this is provided by a work studying the effects of mutating charged amino acids exclusively located on the surface of Bacillus subtilis lipase A (BsLA) [85], an enzyme widely utilized in industrial biocatalysis and thoroughly studied by site-directed mutagenesis. The authors planned to merge eight different mutations on the surface of the protein, all introducing negatively charged amino acids (glutamic and aspartic acids); four of them had been previously identified in a set of twelve substitutions capable of increasing BsLA thermostability [86]; the other four substitutions had been associated to an IL-tolerant phenotype [87]. The obtained octuple variant of BsLA demonstrated a surprisingly incremental effect on both thermal resistance and solubility in ILs. Noteworthy, its specific activity was shown to be comparable to the wild-type enzyme and intermediate between the two control quadruple variants maintaining the four thermo-stabilizing and the four IL-solubilizing mutations, separately (with the former control variant having an almost double activity with respect to the wild-type enzyme).

Efforts in engineering proteins surface charges to produce more useful and applicable biocatalysts have been addressed to different classes of enzymes and have been reviewed elsewhere [88]. Besides the aforementioned example regarding a lipase, we can also mention (i) a research producing negatively supercharged cellulases resistant to lignin and better performing on simulated pretreated lignocellulosic biomass [89] and (ii) a work introducing the profitable properties of halophilic enzymes to the mesophilic bovine carbonic anhydrase II, so producing an extreme halotolerant biocatalyst [90].

Indeed, enzyme stability can be influenced by surface electrostatic forces, especially when considering multimeric enzymes, since surface features modulate PPIs and thus are crucial to either proper dimerization/oligomerization, docking to other proteins, or undesired formation of insoluble aggregates.

To provide hints into protein solubilization strategies, Lawrence et al. investigated the effect of protein net charge on aggregation and precipitation [10]. The introduction of many surface mutations that led to both negative or positive “supercharging” caused a very strong improvement in protein solubility upon thermal and chemical denaturation and function restoring upon refolding. Interestingly, the formation of multimers was not significantly reduced and the physiological function was retained (albeit with protein dependent alterations) [10].

In addition to structure-based mutational attempts to increase protein solubility and purification yields by improved electrostatic features, helpful insights can derive from comparative computational analyses of homologous soluble and insoluble enzymes. A recent example of such an approach is provided by a study in which a dataset of Baeyer-Villiger Mono Oxygenases (BVMOs), enzymes of interest to biocatalysis hence to biotechnology, was investigated by comparative electrostatic distance (ED) analysis. In this study, enzyme clustering by ED was found to be independent of phylogenetic classification, as insoluble enzymes from different phylogenetic BVMO classes clustered altogether rather than with soluble members from the same class, opening the route to in silico predictive protocols for the selection of multiple mutants in rational design projects [91].

Indeed, electrostatic forces are crucial to catalysis itself and proper calculation may provide the basis for engineering. For instance, findings about the electrostatic properties of the active site could be used as a mold for future drug design [20]. Since their role in enzyme catalysis is now recognized, computation of long-ranged electric fields could become a primary design parameter in the discovery of new catalytic materials. In their perspective paper [92], Welborn and co-workers discuss how to use electric fields for computational optimization of the biocatalytic performance of a synthetic enzyme. At the same time, the authors suggest that such computation may represent a unifying descriptor for catalytic design when considering a range of catalysts [92]. An accurate analysis of the electrostatic environmental effects may open new routes toward the rational design and optimization of efficient catalysts; however, when using electric field alignments in the reactive centers of complex catalytic systems, much more predictive capacity is needed for computational design providing results transferable to the experimental field [92]. Electrostatic preorganization has been brought into focus by works studying artificial enzymes to be used in biocatalysis and, in particular, diverse variants of Kemp eliminase, an unnatural enzyme used as benchmark for testing protocols for de novo design [92-94]. Kemp elimination, that is the conversion of benzisoxazoles into salicylonitriles, takes place in solution via a delocalized transition state in which a proton transfer and a bond breaking in a 5-membered ring occur simultaneously; no natural enzyme capable of sustaining such conversion as primary and specific reaction has been identified so far. These studies report on progresses of
computational ground plans toward the achievement of enzymes having efficiencies (only) close to those of natural enzymes. Indeed, the best Kemp eliminase obtained so far, a variant named HG-3.17, arose from 17 rounds of mutagenesis (including directed evolution) and screening from the previous, totally designed variant HG-3 [95]. Hybrid QM/MM molecular dynamics simulations have been used to explain reasons for the limited efficiency of the computationally designed enzyme(s) compared to the one evolved in laboratory. Similarly to the truly natural GNMT enzyme (see subsection 1.2), the analysis revealed the fundamental contribute of electrostatic topography not only in providing an optimal conformational complementarity between the synthetic protein and the substrate but, noticeably, between its active site and the reaction process itself [94].

3.2. Antibodies and nanobodies

The interactions of antibodies (Ab) with their target antigens (Ag) are of particular interest because of their high affinity and specificity and their notable medical and biotechnological relevance. Charged residues directly affect antigen-antibody binding by interacting with other local charges on the protein interactor. Moreover, they are involved in intramolecular interactions that determine secondary structures, in turn indirectly modulating the biophysical properties of an antigen-antibody complex. Such a relevant role in determining binding abilities of proteins justify the frequent targeting of charged amino acids in protein engineering works aiming to modulate antibody affinity and selectivity, often driven by energy calculations and electrostatic potential predictive tools. In a computational design study that led to improving the affinity of different antibodies for their protein targets, the electrostatic term turned out to be a better predictor for enhanced binding than the total calculated free energy [96]. According to the authors, the designed mutations improved binding by either increasing electrostatic interactions following the introduction of a charged residue or by substituting a polar residue with unfavorable desolvation with a hydrophobic one.

The antibody 11K2 underwent a 4.7 fold affinity enhancement for its target (MCP-1) following the introduction of a single mutation involving a charged residue in an in silico engineering work and the enthalpic contribution in the transition state (probably as early as in long-range electrostatic attraction) was suggested.
as the main driving force of the improvement [97]. It is worth noting that 5 out of 12 proposed mutations (all involving charged residues) had an enhancing effect on affinity. Yoshida et al. [98] investigated the effect of surface charge manipulations at the interface between IFN-γR domain D1 and a neutralizing antibody, the complexing of which was abolished or notably reduced by singularly disrupting three electrostatic interactions between the antigen and the antibody and restored by charge-exchanged Ag–Ab pairs. These results showed that electrostatic forces have a pivotal role in PPIs (especially by long-range attraction of partner residues). However, the failure of binding restoration in one case, together with proteins biophysical characterization and molecular dynamics of the different complexes analyzed, confirmed electrostatics crucial impact on protein conformation, in turn further affecting binding events. The authors proposed surface charge conversion as a new strategy for boosting the production of antibodies against weakly immunogenic antigens [98], as depicted in Fig. 3a.

While complementarity determining regions (CDRs) on antibody heavy and light chains variable domains (VH and VL) mostly determine antibody affinity for its target, the sequences of framework regions (FRs) on variable domains and constant domains are main determinants of antibody stability [99]. Although these considerations should be taken into account in protein engineering, some case studies have shown that such a distinction is not exhaustive, especially in electrostatic manipulation.

A single charged mutation (D32H) mapping on the L-CDR1 region of hu3F8 antibody was able to significantly increase its affinity for GD2 ganglioside by altering the surface charge distribution and improved its ADCC, CDC and in vivo anti-tumor activities [100]. Even though a second charge mutation (E1K) mapping on L-FR1 further strengthened hu3F8-GD2 affinity, the double mutant did not show better effector functions than the single D32H mutant [100].

Fukunaga and Tsumoto [101] have demonstrated that besides commonly investigated CDRs, antibodies framework regions could also represent the target for mutations aimed at enhancing antibody binding to its target. Following the introduction of cationic and anionic residues in the FRs of single-chain variable fragments (scFvs) directed toward an acidic and a basic epitope, respectively, they observed higher association rates ($k_{on}$) values by SPR and thus hypothesized that long-range electrostatic attraction might improve antibody-antigen affinity. Some concerns about possible scFv stability reduction upon FR engineering were raised according to the high degree of sequence conservation in this region and experimentally confirmed [101].

The particular influence of long-range electrostatic interactions on the association constant (rather than on the dissociation constant) has also been demonstrated for other protein-protein complexes, e.g. the $k_{on}$ and the affinity of the protein inhibitor BLIP for its target, the β-lactamase TEM1, were improved more than 200-fold (while leaving $k_{off}$ rather unaffected) by modulating BLIP surface charges near the binding interface to increase the electrostatic attraction between the two proteins [102].

Antibody-based therapeutics could benefit from electrostatic manipulation by mechanisms other than increasing affinity for the recognized epitope. The manipulation of few charged residues on Fc fragments has been reported as a feasible way to modulate their interchain interaction, in particular enhancing the formation of desired heterodimers and suppressing side Fc homodimerization in bispecific antibodies production [27,103]. Analogously, the introduction of complementary charged protein tags in the constant domains of a Fab heavy and light chains – in proximity to the interchain disulfide bond – led to electrostatic steering-mediated increase in Fab assembly, thus improving secretion from insect cells [104].

The 10E8 antibody, targeting the membrane-proximal external region (MPER) of the gp41 subunit of the HIV-1 envelope glycoprotein (Env), shows a positively-charged patch on the Fab surface which is suggested to favor the binding to its target by engaging electrostatic interactions with negatively-charged membrane phospholipids [105]. 10E8 neutralizing activity against many HIV-1 isolates was increased by broadening its positive surface patch with the introduction of three solvent-exposed arginine residues [105].

Nanobodies are emerging as potent tools in therapeutics development, diagnosis, and research because of their limited dimensions, easy recombinant production and peculiar binding abilities to difficult-to-target antigens [106]. Electrostatic interactions are gaining a notable interest also in nanobodies engineering.

Recently a structure-based computational method was used to design a single point mutation involving a charged amino acid on the V_{H}NAC1 nanobody, which consequently showed an increase by more than an order of magnitude in affinity for nonamyloid component (NAC) region of human α-synuclein: this improved binding was predicted to stem from electrostatic complementarity optimization at the binding interface between the nanobody and its target [107]. Analogously, Cheng et al. [108] applied an in silico affinity maturation workflow to an anti-CD47 nanobody, obtaining a set of four mutations that improved both target binding and thermostability, combining the introduction of a salt bridge with other different interactions. The last observation suggests that electrostatic engineering could reach better protein optimization if applied together with other design strategies.

The strong binding of a recently developed nanobody to HIV-1 p24 capsid protein was revealed to be due to several interactions, including hydrogen bonds, hydrophobic interactions, and most notably, a salt bridge and the alignment of electrostatic complementarity regions [109]. In particular, the same nanobody was able to bind several p24 subtypes that conserved electrostatic potential homogeneity, while if the broad electrostatic complementarity between antigen and nanobody surfaces was altered by mutations in either of the two proteins, then their binding was diminished [109]. According to this observation, antigenic types classification based on their surface electrostatic potential could be useful in predicting antibodies cross-reactivity against them.

### 3.3. Electrostatic binding to small molecules and nanomaterials

Electrostatic interactions are particularly relevant in small molecules binding to proteins, the former being a substrate, an effector/modulator or an inhibitor. Electrostatic interactions, either transient or stable, together with shape complementarity and hydrophobic interactions, play a key role in defining the enthalpic contribution in ligand recognition. Since electrostatics impact slowly decreases with distance, it allows the establishment of long-range interactions.

Medicinal chemistry studies devoted to developing and optimizing hits and leads that could target the activity of enzymes or receptors rely on the availability of structural data as well as on our proficiency in predicting such interactions where direct experimental evidence is missing [110,111]. Although with different degree of success, a large number of strategies have been recently proposed to evaluate and simulate the impact of electrostatic interactions on this process [80,112–114].

For example, Waldner and co-workers developed an algorithm that was able to predict the correlation between electrostatic features and substrate specificity of nine representative members of the chymotrypsin family of serine proteases [113]. In serine protease, target recognition occurs through a pattern of sub-pockets that build up the active site cleft features and define upstream and downstream amino acids selectivity within the peptide substrate [113].
Through the building and usage of electrostatic molecular interaction fields (by the software GRID; \cite{115}), their method was able to infer electrostatic substrate preferences both at subpockets close to the cleavage site and far from it, correctly predicting similarities and differences within evaluated proteases. The success of this approach, where Van der Waals interactions contribution was minimized on purpose, demonstrated the relevance of electrostatic interactions in the activity and selectivity of chymotrypsin family members.

Another example of a powerful methodology for estimating the electrostatic contribution in view of drug design purposes is offered by the work of Bauer and collaborators \cite{80}. They developed an electrostatic complementarity (EC) calculation and visualization tool that allows evaluating ligand-protein matching at the interface between the two partners and could guide binding and selectivity improvement. Examples from protein families as diverse as kinase inhibitors, glutamate receptors, G protein-coupled receptors and the leukemia drug target Mcl-1, member of bcl-2 family, have been used by the authors to prove the efficacy of the proposed method as well as good matching between predicted electrostatic complementarity and small molecules bioactivity. Rapid identification of suboptimal electrostatic interactions can be used to guide the introduction of changes able to improve binding.

Activities targeted by such approaches could be highly specific, as well as quite promiscuous. The level of promiscuity in substrate recognition by the macromolecule of interest should be taken into account since off-target binding events as well as elusive results might be related to the selectivity of the targeted interactions.

On the other hand, the necessity of broad-spectrum activity inhibitors, as in the case of β-lactamases inhibitors, definitely pose a demand for tackling with families of enzymes with few common features and large variability in their subclasses \cite{116}. A successful example is offered by boronic acid-based inhibitor taniboractam, a molecule currently under clinical development active against a molecule currently under clinical development active against many β-lactams, since it establishes electrostatic interactions with Ser/Thr as cell entry receptors \cite{138}. This Receptor-Binding Domain (RBD) that mediates docking to the host cell by binding sialic acids (SA) as cell entry receptors, plays a key role in infection and sensitivity to vaccines, as it is the major capsidic protein and main viral surface antigen, which mediates attachment and penetration into the host cell; mutations at the HA surface may result in antigenic drift and allow the virus to escape anti-HA Ab neutralization \cite{138}. Mature HA monomers - which form trimers at the viral surface - show a globular head or Receptor-Binding Domain (RBD) that mediates docking to the host cell by binding sialic acids (SA) as cell entry receptors. This mechanism explains why HA is the pivotal player in host specificity \cite{139} and, when mutations result in increased binding to ∼2-6 SA, affinity to the human host is improved \cite{140,142}. In recent years, a comparative functional analysis of AI virus HA was performed by computational prediction of the haemagglutinin RBD structure and surface electrostatics. Finding that electrostatic closeness can group HAs and their RBDs from different H5N1 virus phylogenetic groups suggested that this kind of analysis could provide function-related, rather than taxonomy-related, classification and thus unveil functional fingerprints \cite{143}. This was confirmed by a deeper analysis on H5N1 clades and subclades, which unveiled electrostatic fingerprints related to clades evolution and spreading. In particular, charge redistribution at the RBD surface was found to relate to the branching of still-circulating clades relative to no longer circulating ones, hence being likely involved in antigenic drift events \cite{143}.

### 3.4. Electrostatic fingerprints in the evolution of pandemic viruses

Widespread outbreaks (pandemics) of viruses able to mediate severe respiratory diseases, depending on sporadic “host jump” events \cite{134}, resulted since 1918 in the death of tens of million people worldwide \cite{135}. Especially in the case of influenza, this picture is further complicated by the emergence of novel reassortant viruses, especially where multiple strains and clades co-circulate \cite{136}. These considerations highlight the importance of shedding light on mechanisms underlying viruses evolution and spreading, in order to boost a coordinated global surveillance network by monitoring genetic changes and predicting ‘evolutionary trends’ among emerging viruses for which animal/human host switching has been reported or is likely to occur \cite{137}. Despite an extensive body of literature concerning evolutionary aspects of viral pandemics linked, for example, to antigenic drift, immune escape, interspecies transmission, host specificity shift and low-to-high pathogenicity shift, any “functional” overall explanation and identification of major protein players in viral clade evolution and spreading is still largely missing. Therefore, improving strategies in functional studies about virus variation is dramatically relevant to human health and vaccine efficacy, as well as to animal husbandry.

In Avian influenza (AI) viruses, haemagglutinin (HA) is the central player in infection and sensitivity to vaccines, as it is the major capsidic protein and main viral surface antigen, which mediates attachment and penetration into the host cell; mutations at the HA surface may result in antigenic drift and allow the virus to escape anti-HA Ab neutralization \cite{138}. Mature HA monomers - which form trimers at the viral surface - show a globular head or Receptor-Binding Domain (RBD) that mediates docking to the host cell by binding sialic acids (SA) as cell entry receptors \cite{138}. This mechanism explains why HA is the pivotal player in host specificity \cite{139} and, when mutations result in increased binding to ∼2-6 SA, affinity to the human host is improved \cite{140,142}. In recent years, a comparative functional analysis of AI virus HA was performed by computational prediction of the haemagglutinin RBD structure and surface electrostatics. Finding that electrostatic closeness can group HAs and their RBDs from different H5N1 virus phylogenetic groups suggested that this kind of analysis could provide function-related, rather than taxonomy-related, classification and thus unveil functional fingerprints \cite{143}. This was confirmed by a deeper analysis on H5N1 clades and subclades, which unveiled electrostatic fingerprints related to clades evolution and spreading. In particular, charge redistribution at the RBD surface was found to relate to the branching of still-circulating clades relative to no longer circulating ones, hence being likely involved in antigenic drift events \cite{143}. Further comparative analyses on a different AI virus subtype, H9N2, allowed to confirm that the electro-
static variation by surface charge redistribution is a general fingerprint in AI virus evolution, and a hallmark for AI viruses [144]. Evidence that surface electrostatics is a major player in the evolution and antigenic drift of AI viruses prompted further studies. Specifically, they focused on electrostatic distance (ED)-based grouping of model AI subtypes, either having “pandemic history” (H5 and H7) or a poor story of infections in mammals (H4 and H6), and aimed to infer possible difference in electrostatic clustering and/or in shared fingerprints, among viruses isolated from avian and human/mammalian host. Such a first systematic analysis of the surface electrostatics versus host specificity relationship also provided a deeper analysis of different HA subregions [145]. In particular, clustering by electrostatic closeness of the RBD resulted in groups including mixed avian- and human/mammalian-host viruses and, when the relative distribution among electrostatic groups was considered, H5 and H7 subtypes showed preferential (even if not 100% specific) clustering, as one electrostatic group was fully populated by viruses from human or mammalian host. In contrast, in other groups, viruses from avian hosts represented the major population [145]. Even though the predictive power of electrostatic isocontours for host jump was found to be lower than fingerprints for antigenic drift and clades evolution, the observed preference in clustering suggested electrostatic changes could somehow be involved (likely together with other surface features) in the modulation of host specificity. This discovery elicited deeper analyses by progressively zooming in relevant antigenic epitopes of the RBD. Preferential clustering by electrostatic distance was confirmed by next “zoom in” analyses with progressively smaller HA fragments and focused around the most antigenic 130-loop to 220-loop fragment of the RBD [145]. This evidence is in agreement with another recent work, highlighting the relevance for host specificity of changes at the 130-loop [146]. Furthermore, it is known that decreasing the positive electrostatic charge in the vicinity of RBD epitopes involved in immune escape could also lead to a lowering of the affinity to sialic acid analogs of cell receptors [147]. The aforementioned works suggest that surface electrostatics can modulate, rather than drive, host specificity anyway representing a determinant more important than hydrophathy [145]. Next investigations on changes in surface electrostatics should be combined to the analysis of local changes in e.g. solvent accessible surface area and/or specific linear and conformational motifs, and integrated by docking simulations, for predicting changes in relative affinities to the different types of sialic acid, and thus trends in host jump and pandemic events.

Indeed, electrostatics studies may be of further help to virologists. For instance, diverse electrostatic characteristics at host-pathogen interfaces influence virus pathogenesis [148], and electrostatics is the major determinant in keeping the influenza virus Matrix Protein M1 conformation stable at different pH values [149]. Electrostatic analysis and clustering could have a positive impact on the pharmacological level: as an example, interfering with electrostatics was of help to develop drugs inhibiting HIV-1 fusion [150] and inhibitors for the influenza virus neuraminidase [151].

4. Concluding remarks

Even though a large computational toolbox is currently available for studying role(s) played by electrostatics in regulation of protein life cycle and its interaction, electrostatic features are still a neglected factor in both basic science and applied, biotechnological projects. However, protein electrostatics can deeply influence protein stability, from capacity to properly dimerize/oligomerize or form complexes to undesired aggregation and precipitation. Furthermore, it is crucial to selection of cofactors, substrates, binding to other macromolecules, subcellular compartments, nanomaterials.

Last but not least, an important indication is emerging, not only from electrostatics studies, which concerns the overall structural and protein surface features. In many fields, “functional similarity” is still inferred as directly related to sequence closeness. This is often true, but not always, as different substitutions may dramatically diverge in functional outcome. For instance, several substitutions and thus high divergence depending on the very long, separate evolutionary history of two proteins derived from a common ancestor might suggest a weak functional relationship, especially when compared with another protein showing only a few substitutions. This can be misleading when a high number of substitutions concerns non conserved regions, or anyway all functional motifs are not changed or keep their relevant features. Conversely, two proteins with highly similar sequences might strongly diverge from a functional point of view, when a limited number of substitutions affect residues (even only one) crucial to e.g. catalysis, or binding of a cofactor, a molecular partner etc. In a few words, sequence-based comparison can fail in properly sorting proteins together based on shared functions, and this task can be better achieved by taking advantage of structure based, computational comparison.

In addition to sharing functions based on structural closeness rather than sequence similarity, the analysis of protein surface features is quite relevant to biotechnological project design. Indeed, this review illustrates some examples in which electrostatic features are clearly able to sort proteins in phylogeny-independent clusters where the shared feature is a function, rather than taxonomic classification. Indeed, most protein interactions occur at the protein surface and accessible cavities, i.e. where motifs are exposed and where electrostatics can act as driving force in guidance, specificity and strength in binding events, i.e. ultimately functional events.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by a “DOR” grant from Padua University to F.F., grant Progetto PRIN 2017 JLBSRX and Cariparo Visiting programme 2018 to L.C. and grant Unipd Stars Consolidator FIRMES and PRIN 2017 FSSSHL “RADIUS” to M.Gi. F.V. is supported by a PhD fellowship from the Padua University. The authors have no conflict of interests to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2020.06.029.

References

[1] Zhou H-X, Pang X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. Chem Rev 2018;118:1691–741. https://doi.org/10.1021/acs.chemrev.7b00305
[2] Tsai M-Y, Zheng W, Balamurugan D, Schafer NP, Kim BL, Cheung MS, et al. Electrostatics, structure prediction, and the energy landscapes for protein folding and binding. Protein Sci 2016;25:255–68. https://doi.org/10.1002/pro.2751.
[3] Strickler SS, Gribenko AV, Gribenko AV, Keiffer TR, Tomlinson J, Reihle T, et al. Protein Stability and Surface Electrostatics: A Charged Relationship. Biochemistry 2006;45:2761–6. https://doi.org/10.1021/bi0600143.
[40] Roca M, Messer B, Warshel A. Electrostatic contributions to protein stability and folding energy. FEBS Lett 2007;581:2065–71. https://doi.org/10.1016/j.

[5] Hendich Z, Tidor B. Do salt bridges stabilize proteins?. A continuum electrostatic analysis. Protein Sci 1994;3:211–26. https://doi.org/10.1002/pro.

[46] Der BS, Kluwe C, Miklos AE, Jacak R, Lyskov S, Gray JJ, et al. Alternative Electrostatics in the Ribosomal Tunnel Modulate Chain Elongation Rates. J Mol Biol 2008;384:73–88. https://doi.org/10.1016/j.

[46] Liu W, Bratko D, Prausnitz JM, Blanch HW. Electrostatic Interactions between Peptides and the Molecular Chaperone DnaK. J Phys Chem B 2003;107:11563–9. https://doi.org/10.1021/.

[5] Knight AM, Culviner PH, Kurt-Yilmaz N, Zou T, Ozkan SB, Cavagnero S. Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. J Mol Biol 2017;292:18608–17. https://doi.org/10.1016/j.

[49] Roca M, Messer B, Warshel A. Electrostatic contributions to protein stability and folding energy. FEBS Lett 2007;581:2065–71. https://doi.org/10.1016/j.

[50] Lawrence MS, Phillips KJ, Liu DR. Supercharging Proteins Can Impart Unusual Resilience. J Am Chem Soc 2007;129:10110–2. https://doi.org/10.1021/ja.

[51] Yeung T, Heit B, Dubuisson JF, Fairn GD, Chiu B, Inman R, et al. Contribution of Membrane Targeting Mechanism of Rac to the Phagosome. J Immunol 2005;175:2381–90. https://doi.org/10.110.

[52] Chesner EM, Zuber RB, Butz J, Zaltsman AV, Zuckerman AJ, Stonik AV, et al. Interfacial electrostatics contribute favorably to the conformational stability of ribonuclease T1. Biochemistry 1999;38:13379–87. https://doi.org/10.1021/bi.

[53] Oshima A, Sumimoto K. Vesicle fusion with bilayer lipid membrane by dyshdrophilic refactose and other enzymes arises from electrostatic preorganization, not conformational motions. Proc Natl Acad Sci U S A 2011;108:14115–20. https://doi.org/10.1073/p.

[54] Ghaemi Z, Guzmán I, Gnutt D, Gruebele M. Role of electrostatic and steric stabilization of the carbonium ion in the reaction of glycine N-methyltransferase from Computational Modeling. J Phys Chem B 2015;119:13945–57. https://doi.org/10.1021/

[55] Yuan Y, Zhang C-J, Liu B. A Photoactivatable AIE Polymer for Light-Controlled modification of nanocarriers. Nanomedicine 2015;10:1923–40. https://doi.org/10.2174/

[56] Ghaemi Z, Guzmán I, Gnutt D, Gruebele M. Role of Electrostatics in Protein-RNA Binding: The Global vs the Local Energy Landscape. J Phys Chem B 2017;121:9437–46. https://doi.org/10.1021/

[57] Zhao L, Suarez IP, Gaut D, Rasia RM, Wang J. The key role of electrostatic interactions in the induced folding in RNA recognition by DCL1-A. Phys Chem Chem Phys 2018;20:9376–88. https://doi.org/10.1039/c8cp.

[58] Wang Y, Li H, Xiao Y. 3D-RPC: a web server for 3D-protein structure prediction. Bioinformatics 2018;34:1238–40. https://doi.org/10.1093/b.

[59] Qin S, Zhu H-X. Dissection of the high rate constant for the binding of a ribotoxin to the ribosome. Proc Natl Acad Sci U S A 2006;103:6974–8. https://doi.org/10.1073/p.

[60] Hippenmeier S, Tschopp M, Berghen A, Fesik SW, Gartner J. A photoactivatable AIE dye for imaging proteins in living cells. Angew Chem Int Ed 2019;58:18903–9. https://doi.org/10.1002/a.

[61] Ohshima A, Sumimoto K. Vesicle fusion with blayer lipid membrane controlled by electrostatic interaction. Biochem Biophys Res Commun 2017;1:58–63. https://doi.org/10.1016/j.

[62] Yeung T, Heit B, Dubuisson JF, Fairn GD, Chiu B, Inman R, et al. Contribution of Oshima A, Surpitzo K, Vesicle fusion with blayer lipid membrane controlled by electrostatic interaction. Biochem Biophys Res Commun 2017;1:58–63. https://doi.org/10.1016/j.

[63] McGarvey MAO, Calgolov M. Pivotal Advance: Phospholipids determine net membrane surface charge resulting in differential localization of active Rac1 and Rac2. J Leukoc Biol 2010;87:545–55. https://doi.org/10.118.

[64] Ueyama T, Eto M, Kami K, Tatsuno T, Kobayashi T, Shirai Y, et al. Isomorph-Specific Membrane Targeting Mechanism of Rac during Fc-yR-Mediated Phagocytosis: Positive Charge-Dependent and Independent Targeting Mechanism of Rac to the Phagosome. J Immunol 2005;175:2381–80. https://doi.org/10.404.

[65] Mesmin B, Robbe K, Geny B, Luthey-Schulten Z, Gruebele M. Electrostatics Calculations. J Phys Chem B 2015;119:13945–57. https://doi.org/10.1021/

[66] Yeung T, Gilbert GE, Shi J, Silvius J, Kapus A, Grinstein S. Membrane Phosphatidylserine-binding Site in the Cytosolic Fragment of Clostridium difficile toxin B and other enzymes arises from electrostatic preorganization, not conformational motions. Proc Natl Acad Sci U S A 2011;108:14115–20. https://doi.org/10.1073/p.

[67] Persson BA, Lund M. Association and electrostatic steering of α-lactalbumin–lysozyme heterodimers. Phys Chem Chem Phys 2009;11:3797–81. https://doi.org/10.1039/b.

[68] Carpenter KL, Yablonski K, Hustedt J, Herbots L, Ackermans M, Bullerdijk B, et al. Analysis of amyloid fibril structure using optical tweezers. Proc Natl Acad Sci 2013;110:9028–33. https://doi.org/10.1073/p.

[69] Hillebrand M, Huyse J, Buysse K, Groeneveld AA, De Keersmaecker F, Van Damme P, et al. Inhibition of mannose-mediated fusion of glycoconjugates by a novel AChE inhibitor. J Pharmacol Exp Ther 2017;362:377–85. https://doi.org/10.112.

[70] Polder PG, van der Zee H, de Goede RA, de Groot RA. A Unique Mitochondria-Associated Protein. Trends Cell Biol 2004;14:354–8. https://doi.org/10.1016/j.

[71] Liu W, Bratko D, Prausnitz JM, Blanch HW. Electrostatic Interactions between Peptides and the Molecular Chaperone DnaK. J Phys Chem B 2003;107:11563–9. https://doi.org/10.1021/

[72] Gorham RD, Kieslich CA, Morikis D. Electrostatic Design of Protein–Protein Association Rates. Protein Sci, vol. 340, New Jersey: Humana Press; 2006, p. 235–50. https://doi.org/10.118.

[73] Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM. Electrostatics in the Ribosomal Tunnel Modulate Chain Elongation Rates. J Mol Biol 2008;384:73–88. https://doi.org/10.1016/j.

[74] Warshel A, Sharma PK, Li J, Chen MH, Vance JE. A unique mitochondria-associated protein. Proc Natl Acad Sci U S A 2011;108:15013–4. https://doi.org/10.1073/p.

[75] Krzeminska A, Moliner V, Swither K, Jacak R, Lyskov S, Gray JJ, et al. Alternative Electrostatics in the Ribosomal Tunnel Modulate Chain Elongation Rates. J Mol Biol 2008;384:73–88. https://doi.org/10.1016/j.

[76] Marczynski G, Yokoyama J, Kuwada S, Sakurai Y, Nakamura T. Dynamic and Electrostatic Effects on the Reaction Catalyzed by HIV-1 Protease. J Am Chem Soc 2007;129:10110–2. https://doi.org/10.1021/ja.

[77] Roca M, Messer B, Warshel A. Electrostatic contributions to protein stability and folding energy. FEBS Lett 2007;581:2065–71. https://doi.org/10.1016/j.

[78] Heit B, Chiu B, Inman R, et al. Contribution of phosphatidylserine-binding site in the cytosolic fragment of Clostridium difficile toxin B and other enzymes arises from electrostatic preorganization, not conformational motions. Proc Natl Acad Sci U S A 2011;108:14115–20. https://doi.org/10.1073/p.

[79] Persson BA, Lund M. Association and electrostatic steering of α-lactalbumin–lysozyme heterodimers. Phys Chem Chem Phys 2009;11:3797–81. https://doi.org/10.1039/b.

[80] Yuan Y, Zhang C-J, Liu B. A Photoactivatable AIE Polymer for Light-Controlled Gene Delivery: Concurrent Endolysosomal Escape and DNA Unpacking. F. Vascon et al. / Computational and Structural Biotechnology Journal 18 (2020) 1774–1789. 1787
