Electrostatic Complementarity at Protein/Protein Interfaces

Airlie J. McCoy*, V. Chandana Epa and Peter M. Colman

Calculation of the electrostatic potential of protein–protein complexes has led to the general assertion that protein–protein interfaces display “charge complementarity” and “electrostatic complementarity”. In this study, quantitative measures for these two terms are developed and used to investigate protein–protein interfaces in a rigorous manner. Charge complementarity (CC) was defined using the correlation of charges on nearest neighbour atoms at the interface. All 12 protein–protein interfaces studied had insignificantly small CC values. Therefore, the term charge complementarity is not appropriate for the description of protein–protein interfaces when used in the sense measured by CC. Electrostatic complementarity (EC) was defined using the correlation of surface electrostatic potential at protein–protein interfaces. All twelve protein–protein interfaces studied had significant EC values, and thus the assertion that protein–protein association involves surfaces with complementary electrostatic potential was substantially confirmed. The term electrostatic complementarity can therefore be used to describe protein–protein interfaces when used in the sense measured by EC. Taken together, the results for CC and EC demonstrate the relevance of the long-range effects of charges, as described by the electrostatic potential at the binding interface. The EC value did not partition the complexes by type such as antigen–antibody and proteinase–inhibitor, as measures of the geometrical complementarity at protein–protein interfaces have done. The EC value was also not directly related to the number of salt bridges in the interface, and neutralisation of these salt bridges showed that other charges also contributed significantly to electrostatic complementarity and electrostatic interactions between the proteins. Electrostatic complementarity as defined by EC was extended to investigate the electrostatic similarity at the surface of influenza virus neuraminidase where the epitopes of two monoclonal antibodies, NC10 and NC41, overlap. Although NC10 and NC41 both have quite high values of EC for their interaction with neuraminidase, the similarity in electrostatic potential generated by the two on the overlapping region of the epitopes is insignificant. Thus, it is possible for two antibodies to recognise the electrostatic surface of a protein in dissimilar ways.

Keywords: charge complementarity; continuum electrostatics model; electrostatic complementarity; protein interaction; protein interfaces

*Corresponding author

Present address: A. J. McCoy: Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Abbreviations used: HEL, hen egg white lysozyme; D1.3, anti-HEL antibody D1.3; hGH, human growth hormone; hGHR, human growth hormone receptor; HyHEL10, anti-HEL antibody HyHEL10; HyHEL5, anti-HEL antibody HyHEL5; NA (N9), influenza virus neuraminidase (subtype 9); NC10, anti-NA antibody NC10; NC41, anti-NA antibody NC41; Fv, variable domain of antibody; Fab, antigen-binding fragment of antibody; OMKTY3, turkey ovomucoid third domain; V_{HV} variable domain of the heavy chain of antibody.
Introduction

The nature of the electrostatic interaction between proteins has long been a subject of study. Elementary analysis has involved simply counting the number of charged residues and salt bridges at protein–protein interfaces (Janin & Chothia, 1990; Jones & Thornton, 1995). With the development of the continuum electrostatics model for proteins, it has become routine to calculate and display the electrostatic potential of protein structures by numerically solving the Poisson-Boltzmann equation for the protein–solvent system (as implemented in the program DelPhi; Gilson & Honig, 1988; Nicholls & Honig, 1991). The electrostatic potential is generally visualised by colour coding equipotential contours in the space around the protein. Such visualisations have been important in gaining insight into the docking of proteins and their small molecule ligands (for example, superoxide and copper, zinc superoxide dismutase; Getzoff et al., 1983) and predicting sites of protein–protein association (for example, nerve growth factor; McDonald et al., 1991). Where the structure of a protein–protein complex has been determined, continuum electrostatics models can be used to visualise the electrostatic potential at the surface buried in the interface and to calculate the contribution to the free energy of binding made by the electrostatic union of the proteins (Gilson & Honig, 1988).

Various analyses of the electrostatic nature of protein–protein interfaces using these techniques has led to the general assertion that proteins and protein surfaces that interact with one another have "charge complementarity" (for example, Novotny & Sharp, 1992; Roberts et al., 1991) or "electrostatic complementarity" (for example, Braden & Poljak, 1995; Demchuk et al., 1994; Hendsch & Tidor, 1994; Karshikov et al., 1992; Lescar et al., 1995; Novotny & Sharp, 1992), and that this property of protein–protein interaction is important for defining specificity. However, to date there has not been a rigorous study of the complementarity of the charges and electrostatic potential at protein–protein interfaces. The aim of this study is to define quantitative measures for charge complementarity and electrostatic complementarity in agreement with their qualitative descriptions as measurements of the complementarity of the charges and electrostatic potential, respectively, at the protein–protein interface itself. These measures for charge complementarity and electrostatic complementarity are then calculated for a variety of protein–protein complexes, and conclusions drawn about the validity of the measurements and the nature of protein–protein interactions. We examine how electrostatic complementarity is affected by the various parameters used for modelling the electrostatic potential with the finite difference Poisson-Boltzmann method, in order to select the most appropriate parameters for the comparative study. We show that while the electrostatic potential at protein–protein interfaces can exhibit quite significant electrostatic complementarity, none of the protein complexes selected shows any significant charge complementarity. Salt bridges across the interface can play an important role in determining the magnitude of complementarity, but because of the effect of all the other charges there can be significant electrostatic complementarity even when the salt bridges are computationally neutralised. Finally, we examine the similarity of the electrostatic potential generated by antibodies NC10 and NC41 at the surface of influenza virus neuraminidase protein (NA) where their epitopes overlap and find that it is insignificant, even though each antibody has high electrostatic complementary with NA at the overlapping epitopes.

Theory

The measures of charge and electrostatic complementarity developed here are based on the correlation coefficients used by Chau & Dean (1994) to analyse complementarity at the interfaces between proteins and their small molecule ligands (for example, retinol-binding protein and retinol). The study also uses the interaction energy, \( \Delta G_{\text{int}} \) and the Coulombic energy, \( \Delta G_{\text{Coul}} \), which are components of the total electrostatic free energy \( \Delta G_{\text{elec}} \) calculated by two different thermodynamic pathways for protein association as modelled by continuum electrostatics (Gilson & Honig, 1988).

Measure of charge complementarity: CC

First, all nearest neighbour contacts made between the proteins in the complex are determined. Where an atom is in contact with more than one atom in the other protein, each contact is counted separately. Charges are then assigned to the atoms in contact, and the correlation coefficient \( r \) between the lists of charges calculated. The charge complementarity (CC) is then defined as:

\[
CC = -r
\]

(1)

The more positive the value of CC, the more complementary are the charges on the surface of the interface, and the more negative the CC value, the more similar is the distribution of charges on the interface. Pearson’s correlation test or Spearman’s rank correlation test (see Materials and Methods) are used for calculating the correlation. Thus, two CC values are calculated:

\[
CC^P = -r^P
\]

(2)

\[
CC^S = -r^S
\]

(3)

where the superscript indicates the type of correlation calculated: \( P \) refers to Pearson’s correlation coefficient and \( S \) refers to Spearman’s rank correlation coefficient.
As the charge of each atom is not given by X-ray diffraction data, the charges must be assigned from a parameter set. Formal and PARSE (Sitkoff et al., 1994) charge parameter sets were used in this study. The Formal charge set only distributes partial charges to the peptide main chain (N = -0.35; H = +0.25; Cα = +0.1; C = +0.55; O = -0.55), formally charged residues (Lys Nε = +1.0; Arg Nε1,ε2 = +0.5; Glu Oε1,ε2 = -0.5; Asp Oε1,ε2 = -0.5), and the amino (N = +1.0) and carboxy (O = -1.0) termini. The PARSE charges and radii have been parameterized to reproduce the experimental solvation energies of amino acids when computed with DelPhi (Sitkoff et al., 1994). Partial charges are assigned to atoms in all amino acid side-chains even if the net charge on the residue is zero.

Measure of electrostatic complementarity: $EC$

The molecular surface on each of the proteins (proteins 1 and 2) buried upon complex formation is calculated to give two interface surfaces (surfaces 1 and 2). For the electrostatic fields generated by proteins 1 and 2 of the complex individually, the values of each of the electrostatic potentials at each point of the two interface surfaces 1 and 2 are determined, generating four lists of electrostatic potential (one for each combination of electrostatic field and surface). We use the finite difference Poisson-Boltzmann method (Gilson & Honig, 1988), as implemented in DelPhi, to compute this electrostatic potential. In this continuum electrostatic model, the interior of the molecule is considered to have a low dielectric constant while the surrounding solvent is taken to be a continuum of high dielectric constant.

Correlations are next calculated between the electrostatic potential generated by proteins 1 and 2 on the buried molecular surface of protein 1 ($r^1$) and also between the electrostatic potentials generated by proteins 1 and 2 on the buried molecular surface of protein 2 ($r^2$). Electrostatic complementarity ($EC$) is then defined as:

$$EC = -\langle r^1 + r^2 \rangle / 2$$

$EC$ is thus positive when the electrostatic potentials on the surface are complementary and negative when the electrostatic potentials on the surface are similar.

The electrostatic potential depends on the position of the dielectric boundaries used to generate the electrostatic field. The electrostatic potential generated by either fully solvated proteins (Figure 1a)), or proteins partially desolvated by the volume of the other protein in the complex (Figure 1b)), can be used. Thus, the following two $EC$ values are calculated:

$$EC_{FS} = -\langle r^1_{FS} + r^2_{FS} \rangle / 2$$

$$EC_{PS} = -\langle r^1_{PS} + r^2_{PS} \rangle / 2$$

where the subscript indicates the type of electrostatic potential: $FS$ refers to electrostatic potential generated by fully solvated proteins, and $PS$ refers to electrostatic potential generated by partially solvated proteins.

The value of $EC$ also depends on the type of correlation calculated, Pearson’s correlation or Spearman’s rank correlation (see Materials and Methods). Thus, when combined with the two types of electrostatic potential, four $EC$ values are calculated:

$$EC_{FS}^p = -\langle r^1_{FS} + r^2_{FS} \rangle / 2$$

$$EC_{FS}^s = -\langle r^1_{FS} + r^2_{FS} \rangle / 2$$

$$EC_{PS}^p = -\langle r^1_{PS} + r^2_{PS} \rangle / 2$$

$$EC_{PS}^s = -\langle r^1_{PS} + r^2_{PS} \rangle / 2$$

where the superscript indicates the type of correlation calculated: $P$ refers to Pearson’s correlation, and $S$ refers to Spearman’s rank correlation.

Electrostatic energies

$\Delta G_{\text{int}}$ is part of a thermodynamic pathway in which $\Delta G_{\text{elec}}$ is divided into partial desolvation and interaction components (Gilson & Honig, 1988). In the first step, each molecule is partially desolvated by removing the high dielectric medium (i.e. the solvent) from the region that the other molecule occupies after binding and replacing it with the low dielectric medium (i.e. the uncharged protein). The charge–solvent interaction energies of this step are $\Delta G_{\text{protein}}^{\text{1}}$ and $\Delta G_{\text{protein}}^{\text{2}}$. In the second step, the interaction energy, $\Delta G_{\text{int}}$, between the proteins in the presence of solvent is calculated by charging the low dielectric cavity now present:

$$\Delta G_{\text{int}} = \Sigma q_i \Phi_i$$

where $q_i$ are the newly created charges, and the potential $\Phi_i$ at $q_i$ is generated by the already charged, partially desolvated molecule. The value of $\Delta G_{\text{int}}$ is the same regardless of which protein is initially charged, so that:

$$\Delta G_{\text{elec}} = \Delta G_{\text{protein}}^{\text{1}} + \Delta G_{\text{protein}}^{\text{2}} + \Delta G_{\text{int}}$$

$\Delta G_{\text{Coul}}$ is part of a thermodynamic pathway in which $\Delta G_{\text{elec}}$ is divided into full desolvation and Coulombic energies (Gilson & Honig, 1988). First, each protein is fully desolvated by removing the high dielectric medium (i.e. the solvent) and replacing it with the low dielectric medium (i.e. same as that of the protein), giving the desolvation energy of each protein, $\Delta G_{\text{protein}}^{\text{1}}$ and $\Delta G_{\text{protein}}^{\text{2}}$. Secondly, the Coulombic energy, $\Delta G_{\text{Coul}}$, between the proteins is calculated as the difference between the energy needed to bring all charges from infinity to the positions defined by the structure of the complex, and the sum of the
energies required to bring the charges of each protein individually from infinity to the same location, in the dielectric medium of the protein. Finally, the solvation energy of the complex, $\Delta G_{\text{solv}}^{\text{protein 1+2}}$, is calculated, so that:

$$\Delta G_{\text{elec}} = -\Delta G_{\text{solv}}^{\text{protein 1}} - \Delta G_{\text{solv}}^{\text{protein 2}} + \Delta G_{\text{Coul}} + \Delta G_{\text{solv}}^{\text{protein 1+2}}$$  \hspace{1cm} (13)

Results and Discussion

Charge complementarity

The calculation of CC requires the identification of atoms neighbouring one another in the interface. The distance chosen as the cut-off for including nearby atoms as neighbours is somewhat arbitrary because the distance over which charge interactions occur is not fixed: the electrostatic potential between two charges approaches zero asymptoti-
involved in "ion pairs" within proteins (Barlow & Thornton, 1983). The distance often taken as the maximum length of separation of charged groups involved in ion pairs within proteins (Barlow & Thornton, 1983), 3.5 Å, the distance often taken as the maximum length of a hydrogen bond (Baker & Hubbard, 1984; Rashin & Honig, 1984), and 4.0 Å the distance often taken as the maximum separation of charged groups involved in "ion pairs" within proteins (Barlow & Thornton, 1983).

Charge complementarity was calculated at each of a variety of protein–protein interfaces chosen from the PDB, using both the Formal and PARSE charge parameter sets. Table 1 shows the calculated CC values for these interfaces using a cut-off distance of 3.5 Å. It can be seen that none of the complexes has significant CC_p or CC_s values for either of the charge parameter sets. Similarly insignificant values are obtained at cut-off distances of 3.0 Å and 4.0 Å (results not shown). Therefore, protein–protein interfaces do not exhibit charge complementarity as measured by CC. A similar conclusion was drawn by Chau & Dean (1994) for protein–small molecule ligand interfaces.

For all but one of the 24 proteins involved in the 12 complexes considered, the sum of the PARSE assigned charges on atoms making an interface contact within a cut-off distance of 3.0 Å is between 0 and ~7.5. The exception is OMKTY3 in the chymotrypsin/OMKTY3 complex (+4.2). The presence of negative charge on the solvent-exposed surface of proteins has been noted (Novotny & Sharp, 1992) and attributed to the high partial charge of carbonyl oxygens (q = ~0.55), which are commonly oriented towards the surface of the protein. The propensity for the interface surfaces to be negatively charged decreases with increasing cut-off distance: when the cut-off is 3.5 Å, four interfaces carry a net positive charge, and when it is 4.0 Å, eight interfaces carry a net positive charge. Most of the trend towards positive values is due to a reduction in the dominance of the negatively charged carbonyl oxygen atoms by the inclusion of the positively charged backbone carbon atoms to which they are bonded, within the larger cut-off distances.

<table>
<thead>
<tr>
<th>PDB</th>
<th>Protein 1</th>
<th>Protein 2</th>
<th>CC_p</th>
<th>CC_s</th>
<th>CC_p</th>
<th>CC_s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1CHO</td>
<td>Chymotrypsin</td>
<td>OMKTY3</td>
<td>-0.06</td>
<td>-0.05</td>
<td>-0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>1FDL</td>
<td>HEL</td>
<td>D1.3</td>
<td>-0.08</td>
<td>-0.04</td>
<td>-0.07</td>
<td>-0.06</td>
</tr>
<tr>
<td>1NCA</td>
<td>NA</td>
<td>NC41</td>
<td>0.04</td>
<td>0.06</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>1NMB</td>
<td>NA</td>
<td>NC10</td>
<td>0.01</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>2HFL</td>
<td>HEL</td>
<td>HyHEL5</td>
<td>0.01</td>
<td>0.00</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>2PTC</td>
<td>Trypsin</td>
<td>BPTI</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>2SEC</td>
<td>Subtilisin</td>
<td>Egin c</td>
<td>-0.08</td>
<td>-0.04</td>
<td>-0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td>2SNF</td>
<td>Subtilisin</td>
<td>CI-2</td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.07</td>
<td>-0.08</td>
</tr>
<tr>
<td>3HFM</td>
<td>HEL</td>
<td>HyHEL10</td>
<td>0.01</td>
<td>0.04</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>3HHR</td>
<td>hGH</td>
<td>hGH4_1</td>
<td>0.02</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>3HHR</td>
<td>hGH-hGH4_1</td>
<td>hGH4_2</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>4HTC</td>
<td>Thrombin</td>
<td>Hirudin</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 2. Electrostatic complementarity at selected protein–protein interfaces

<table>
<thead>
<tr>
<th>PDB</th>
<th>Protein 1</th>
<th>Protein 2</th>
<th>$EC_{PS}$</th>
<th>$EC_{PS}$</th>
<th>$EC_{PS}$</th>
<th>$EC_{PS}$</th>
<th>$\Delta G_{int}$</th>
<th>$\Delta G_{Coul}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1CHO</td>
<td>Chymotrypsin</td>
<td>OMKTY3</td>
<td>0.16 ± 3</td>
<td>0.53 ± 5</td>
<td>0.52 ± 1</td>
<td>0.47 ± 3</td>
<td>−41</td>
<td>−95</td>
</tr>
<tr>
<td>1FDL</td>
<td>HEL</td>
<td>DL3 Fab</td>
<td>0.14 ± 2</td>
<td>0.39 ± 4</td>
<td>0.57 ± 1</td>
<td>0.65 ± 3</td>
<td>−36</td>
<td>84</td>
</tr>
<tr>
<td>1FDL</td>
<td>HEL</td>
<td>DL3 Fv</td>
<td>0.11 ± 1</td>
<td>0.40 ± 3</td>
<td>0.57 ± 0</td>
<td>0.65 ± 3</td>
<td>−36</td>
<td>112</td>
</tr>
<tr>
<td>1NCA</td>
<td>NA</td>
<td>NC41 Fab</td>
<td>0.21 ± 3</td>
<td>0.44 ± 6</td>
<td>0.71 ± 0</td>
<td>0.61 ± 2</td>
<td>−102</td>
<td>−176</td>
</tr>
<tr>
<td>1NCA</td>
<td>NA</td>
<td>NC41 Fv</td>
<td>0.19 ± 4</td>
<td>0.44 ± 6</td>
<td>0.71 ± 1</td>
<td>0.62 ± 2</td>
<td>−100</td>
<td>−169</td>
</tr>
<tr>
<td>1NMB</td>
<td>NA</td>
<td>NC10</td>
<td>0.15 ± 3</td>
<td>0.28 ± 3</td>
<td>0.61 ± 2</td>
<td>0.54 ± 3</td>
<td>−52</td>
<td>−86</td>
</tr>
<tr>
<td>2HFL</td>
<td>HEL</td>
<td>HyHEL5 Fab</td>
<td>0.22 ± 3</td>
<td>0.59 ± 7</td>
<td>0.76 ± 2</td>
<td>0.69 ± 6</td>
<td>−87</td>
<td>−195</td>
</tr>
<tr>
<td>2HFL</td>
<td>HEL</td>
<td>HyHEL5 Fv</td>
<td>0.20 ± 4</td>
<td>0.61 ± 6</td>
<td>0.73 ± 2</td>
<td>0.69 ± 6</td>
<td>−81</td>
<td>−209</td>
</tr>
<tr>
<td>2PTC</td>
<td>Trypsin</td>
<td>BPTI</td>
<td>0.07 ± 5</td>
<td>0.13 ± 6</td>
<td>0.28 ± 3</td>
<td>0.40 ± 2</td>
<td>−55</td>
<td>206</td>
</tr>
<tr>
<td>2SEC</td>
<td>Subtilisin</td>
<td>Eglin c</td>
<td>0.15 ± 6</td>
<td>0.67 ± 0</td>
<td>0.56 ± 4</td>
<td>0.68 ± 1</td>
<td>−49</td>
<td>−31</td>
</tr>
<tr>
<td>2SNJ</td>
<td>Subtilisin</td>
<td>Cl-2</td>
<td>0.11 ± 0</td>
<td>0.51 ± 2</td>
<td>0.43 ± 2</td>
<td>0.54 ± 0</td>
<td>−38</td>
<td>−68</td>
</tr>
<tr>
<td>3HFH</td>
<td>HEL</td>
<td>HyHEL10 Fab</td>
<td>0.08 ± 4</td>
<td>0.22 ± 6</td>
<td>0.27 ± 4</td>
<td>0.42 ± 7</td>
<td>−33</td>
<td>−169</td>
</tr>
<tr>
<td>3HFH</td>
<td>HEL</td>
<td>HyHEL10 Fv</td>
<td>0.07 ± 3</td>
<td>0.22 ± 6</td>
<td>0.28 ± 2</td>
<td>0.44 ± 4</td>
<td>−33</td>
<td>−152</td>
</tr>
<tr>
<td>3HHR</td>
<td>hGH</td>
<td>hGH$_{H_4}$</td>
<td>0.20 ± 1</td>
<td>0.61 ± 6</td>
<td>0.66 ± 0</td>
<td>0.65 ± 3</td>
<td>−71</td>
<td>−55</td>
</tr>
<tr>
<td>3HHR</td>
<td>hGH</td>
<td>hGH$_{H_2}$</td>
<td>0.21 ± 8</td>
<td>0.37 ± 1</td>
<td>0.62 ± 2</td>
<td>0.51 ± 1</td>
<td>−87</td>
<td>105</td>
</tr>
<tr>
<td>4HTC</td>
<td>Hirudin</td>
<td>Thrombin</td>
<td>0.21 ± 0</td>
<td>0.70 ± 0</td>
<td>0.61 ± 2</td>
<td>0.76 ± 2</td>
<td>−107</td>
<td>−579</td>
</tr>
</tbody>
</table>

Values are given as $EC_{\pm} \Delta$, where $EC_{+}$ and $EC_{-}$ are correlations $r^2$ and $r^2$; $\Delta$ given in the last decimal place. Energies in kcal mol$^{-1}$ at 25°C.

* Fujinaga et al. (1987). $K_a = 1.8 \times 10^{11}$ M$^{-1}$ (Fujinaga et al. (1987).

* Bhat et al. (1990). $K_a = 2.7 \times 10^{10}$ M$^{-1}$ (Bhat et al. (1994).

* Tulip et al. (1992). Where the Fab was truncated to an Fv fragment, residues 1 to 109 of the variable domain light chain and residues 1 to 113 of the variable domain heavy chain were included. $K_a = 7.0 \times 10^{7}$ M$^{-1}$ (L.C. Gruen & A. A. Kortt, unpublished results).

* Malby et al. (1994). The constant domain of the Fab was not modelled. $K_a = 1.1 \times 10^{10}$ M$^{-1}$ (L.C. Gruen & A. A. Kortt, unpublished results).

* Sherrill et al. (1987). The non-standard amino acid at the amino terminus of the heavy chain of the antibody was deleted as there were no charge parameters for this residue. Where the Fab was truncated to an Fv fragment, residues 1 to 109 of the variable domain light chain and residues 1 to 113 of the variable domain heavy chain were included. $K_a = 4.3 \times 10^{8}$ M$^{-1}$ (Janin, 1995).

* Huber et al. (1974).

* McPhalen & James, 1988. $K_a = 4.0 \times 10^{8}$ M$^{-1}$ (Janin, 1995).

* Padlan et al. (1989). Where the Fab was truncated to an Fv fragment, residues 1 to 109 of the variable domain light chain and residues 1 to 113 of the variable domain heavy chain were included. $K_a = 1.5 \times 10^{8}$ M$^{-1}$ (Padlan et al. 1989).

* de Vos et al. (1992). The co-ordinates were used to represent two complexes: the binding of hGH to the monomeric hGH$_{1}$ (hGH$_{H_4}$), and the binding of the second hGH$_{1}$ molecule (hGH$_{H_2}$) to the preformed hGH–hGH$_{1}$ complex. These complexes reflect the sequential mode of binding of hGH to its receptor. $K_a = 2.5 \times 10^{9}$ M$^{-1}$ for the hGH$_{H_1}$–hGH complex (Cunningham et al., 1991).

* Rydel et al. (1991). $K_a = 4.3 \times 10^{12}$ M$^{-1}$ (Braun et al., 1988).

Electrostatic complementarity

EC was calculated for a selection of proteinase–inhibitor, ligand–receptor and antigen–antibody complexes (see Table 2). $\Delta G_{int}$ and $\Delta G_{Coul}$ were also calculated for each of the protein–protein complexes for comparison with EC. $\Delta G_{dec}$ was not calculated because it includes the (unfavourable) desolvation energy of each protein, which is a property of the interaction of the protein with solvent rather than the interaction between the proteins.

All EC values shown in Table 2 are significantly positive. It thus appears that protein–protein interfaces have anti-correlated (complementary) surface electrostatic potentials. Chau & Dean (1994) drew a similar conclusion for protein-small molecule ligand interfaces. For each complex, $EC_{PS}$ values are much lower than $EC_{PS}$ values. The $EC_{PS}$ and $EC_{PS}$ values are highly correlated with each other so that the ordering of complexes by either parameter is similar. The patterns of electrostatic potential on the buried molecular surfaces for the complexes with the highest and lowest $EC_{PS}$ values (HEL–HyHEL5 Fab and HEL–HyHEL10 Fab) are shown in Figure 2.

Values of EC for the complexes between antigens and Fab fragments are seen to be very similar to the EC value for the complex between the antigens and Fv fragments. This implies that charges and dielectric boundaries over 20 Å from the interface do not significantly affect the electrostatic potential at the interface.

The EC values shown in Table 2 do not partition the complexes by type (proteinase–inhibitor, antibody–antigen or ligand–receptor). This is in contrast to the partitioning of antigen–antibody complexes from other protein–protein complexes by shape complementarity coefficient, $S_C$ (Lawrence & Colman, 1993), and gap volume index (Jones & Thornton, 1995). The lower shape complementarity of antigen–antibody interfaces has been attributed to the differences in evolutionary history between these complexes and other protein–protein complexes (Lawrence & Colman, 1993). The different evolutionary histories of the complexes do not manifest themselves in the electrostatic complementarity of the systems.

It is important to note that EC concerns the electrostatic potential at the interface and as such is not directly related to either $\Delta G_{int}$ or $\Delta G_{Coul}$. EC is high when the peaks in the values of the electrostatic potential generated by one protein match the
Figure 2. The electrostatic potential on the molecular surfaces buried in the interface between the two proteins of the complex. The labelling of the surfaces and type of electrostatic potential used follows the nomenclature introduced in Figure 1. The views of surfaces in each interface are related by a 180° rotation about an axis oriented north–south on the page. For the purposes of the Figures, buried molecular surfaces were calculated using GRASP (written by A. Nicholls & B. Honig), although for the calculation of EC described in the text, the buried molecular surfaces were calculated using the MS program suite. Electrostatic potential was calculated using DelPhi (as implemented in GRASP) as for the calculation of EC described in the text, except that for the Figures, the grid size was limited to a box of dimensions 65 × 65 × 65. Correlation coefficients for the calculation of EC were calculated between the sets of electrostatic potential illustrated across the page in each of the panels. The most negative electrostatic potential throughout all space for each electrostatic field is indicated in red and the most positive electrostatic potential throughout all space for each electrostatic field is indicated in blue. (a) Electrostatic potential on the molecular surfaces buried at the interface between HEL (protein 1) and antibody HyHEL5 (protein 2); PDB entry 2HFL. Electrostatic potential generated by fully solvated proteins. Left column: FS₁, range −15 to 28 kcal mol⁻¹ at 25°C. Right column: FS₂, range −27 to 18 kcal mol⁻¹ at 25°C. Top row: molecular surface of HEL buried by HyHEL5. Bottom row: molecular surface of HyHEL5 buried by HEL. (b) Electrostatic potential on the molecular surfaces buried at the interface between HEL (protein 1) and antibody HyHEL5 (protein 2); PDB entry 2HFL. Electrostatic potential generated by partially solvated proteins. Left column: PS₁, range −15 to 40 kcal mol⁻¹ at 25°C. Right column: PS₂, range −84 to 16 kcal mol⁻¹ at 25°C. Top row: molecular surface of HEL buried by HyHEL5. Bottom row: molecular surface of HyHEL5 buried by HEL. (c) Electrostatic potential on the molecular surfaces buried at the interface between HEL (protein 1) and antibody HyHEL10 (protein 2), PDB entry 3HFM. Electrostatic potential generated by fully solvated proteins. Left column: FS₁, range −22 to 22 kcal mol⁻¹ at 25°C. Right column: FS₂, range −27 to 16 kcal mol⁻¹ at 25°C. Top row: molecular surface of HEL buried by HyHEL10. Bottom row: molecular surface of HyHEL10 buried by HEL. (d) Electrostatic potential on the molecular surfaces buried at the interface between HEL (protein 1) and antibody HyHEL10 (protein 2); PDB entry 3HFM. Electrostatic potential generated by partially solvated proteins. Left column: PS₁, range −47 to 58 kcal mol⁻¹ at 25°C. Right column: PS₂, range −38 to 16 kcal mol⁻¹ at 25°C. Top row: molecular surface of HEL buried by HyHEL10. Bottom row: molecular surface of HyHEL10 buried by HEL.
Table 3. Variation of $EC$, $\Delta G_{int}$ and $\Delta G_{Coul}$ with the parameters of the model for complex NA/NC10, PDB entry 1NMB (Malby et al., 1994)

<table>
<thead>
<tr>
<th>Parameter variation</th>
<th>$EC_1^{\beta}$</th>
<th>$EC_2^{\beta}$</th>
<th>$EC_3^{\beta}$</th>
<th>$EC_7^{\beta}$</th>
<th>$\Delta G_{int}$</th>
<th>$\Delta G_{Coul}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference$^a$</td>
<td>0.15 ± 3</td>
<td>0.28 ± 3</td>
<td>0.61 ± 2</td>
<td>0.54 ± 3</td>
<td>−52</td>
<td>−86</td>
</tr>
<tr>
<td>Sampling of buried molecular surface</td>
<td>0.14 ± 3</td>
<td>0.29 ± 2</td>
<td>0.61 ± 1</td>
<td>0.54 ± 1</td>
<td>−52</td>
<td>−86</td>
</tr>
<tr>
<td>increased from 4 grids/$\AA^2$ to 16 grids/$\AA^2$</td>
<td>0.13 ± 1</td>
<td>0.29 ± 5</td>
<td>0.61 ± 3</td>
<td>0.53 ± 4</td>
<td>−52</td>
<td>−86</td>
</tr>
<tr>
<td>M5 surface probe radius increased from 1.4 Å to 1.7 Å</td>
<td>0.17 ± 3</td>
<td>0.28 ± 1</td>
<td>0.58 ± 3</td>
<td>0.50 ± 1</td>
<td>−56</td>
<td>−83</td>
</tr>
<tr>
<td>DelPhi grid decreased from 201 × 201 × 201 to 65 × 65 × 65 points/side</td>
<td>0.15 ± 3</td>
<td>0.29 ± 3</td>
<td>0.59 ± 2</td>
<td>0.54 ± 1</td>
<td>−59</td>
<td>−86</td>
</tr>
<tr>
<td>DelPhi surface probe radius increased from 1.4 Å to 1.7 Å</td>
<td>0.12 ± 3</td>
<td>0.27 ± 2</td>
<td>0.57 ± 3</td>
<td>0.50 ± 4</td>
<td>−47</td>
<td>−78</td>
</tr>
<tr>
<td>Hydrogen positions changed to those not minimised in the electrostatic field</td>
<td>0.19 ± 2</td>
<td>0.29 ± 3</td>
<td>0.60 ± 2</td>
<td>0.54 ± 3</td>
<td>−27</td>
<td>-43</td>
</tr>
<tr>
<td>Dielectric used to model the interior of protein increased from 2 to 4</td>
<td>0.10 ± 1</td>
<td>0.25 ± 3</td>
<td>0.36 ± 5</td>
<td>0.41 ± 1</td>
<td>NC$^b$</td>
<td>+37</td>
</tr>
</tbody>
</table>

Values given as $EC \pm \Delta$, where $EC+\Delta$ and $EC-\Delta$ are correlations $r^1$ and $r^2$; $\Delta$ given in the last decimal place. Energy in kcal mol$^{-1}$ at 25°C.

$^a$ Parameter values described in Materials and Methods.

$^b$ NC, energy had not converged.

Electrostatic Complementarity

Throughs in the electrostatic potential generated by the other protein, irrespective of the absolute values of the electrostatic potential. $\Delta G_{int}$ is calculated using the electrostatic potential throughout the space of the docked protein and is dependent on the absolute values of the electrostatic potential at each of the atoms. Despite these fundamental differences, $EC$ and $\Delta G_{int}$ are correlated. This arises because the distribution of electrostatic potential for the protein–protein interfaces has a mean of zero and approximately the same standard deviation, so that the distribution of absolute values of electrostatic potential is similar for all complexes. $\Delta G_{Coul}$ is not correlated with $EC$, but is large and negative when the proteins carried large net charges of opposite sign, and is correlated with the product of the net charges on the two proteins. It appears that the complementarity of the surface electrostatic potential at the site of interaction of two proteins is a feature of the docking even when $\Delta G_{Coul}$ is positive (i.e. is unfavourable). In particular $EC$, $\Delta G_{int}$ and $\Delta G_{Coul}$ are not necessarily correlated with the affinity of interaction, since total $\Delta G$ of binding includes contributions from other enthalpic and entropic terms.

Values of the four measures of $EC$ ($EC_1^{\beta}$, $EC_2^{\beta}$, $EC_3^{\beta}$, and $EC_7^{\beta}$), are more stable under changes in the electrostatic parameters of the model than are the electrostatic energies $\Delta G_{int}$ and $\Delta G_{Coul}$ (Table 3; see Materials and Methods for a discussion of the dependence of electrostatic complementarity on various parameters used in the model). In particular, $EC_1^{\beta}$ and $EC_2^{\beta}$ (being correlations) are much more stable than $\Delta G_{int}$ and $\Delta G_{Coul}$ (being sums) to changes in the probe radius used for generating the dielectric boundary and the value of the interior dielectric constant. $EC$ is thus a robust measure of electrostatic complementarity.

Salt bridges

Previous qualitative studies of electrostatic complementarity have been mainly concerned with the presence or absence of salt bridges in the interface of protein–protein complexes. For example, the HEL–HyHEL5 complex has been regarded as a complex with high electrostatic complementarity because there are three salt bridges in the interface, while the HEL–HyHEL10 and HEL–D1.3 complexes have been regarded as having less electrostatic complementarity as they have only one weak salt bridge, and no salt bridges, respectively (Slagle et al., 1994). However, the calculations of $EC$, and indeed $\Delta G_{int}$ and $\Delta G_{Coul}$ take into account all the charges, and not just the net charge on amino acids forming salt bridges in the interface, nor even just all the surface charges. Thus, the relative values of $EC$ for the HEL–HyHEL5, HEL–HyHEL10 and HEL–D1.3 complexes (see Table 2) are not proportional to the number of salt bridges, although the HEL–HyHEL5 complex does have the highest $EC$ value of the three.

In order to determine the influence of salt bridges, $EC$ was recalculated after neutralising the salt bridges at the interface of those protein–protein complexes that have intermolecular salt bridges, i.e. the side-chains involved in the salt bridges were assigned the partial charges of the uncharged amino acid as given in the PARSE charge parameter set. These results are shown in Table 4. The changes in the values of $EC$ vary: for the NA–NC10, NA–NC41, HEL–HyHEL5 and hGH–hGH$_2$ complexes, $EC$ is markedly reduced; for the thrombin–hirudin complex, $EC$ is somewhat reduced; for the trypsin–BPTI, HEL–HyHEL10 and hGH/hGH$_2$–hGH$_2$ complexes, $EC$ is only slightly reduced. Similar variation in the
amount of increase in values of $\Delta G_{\text{int}}$ (towards less negative values) is observed. These variations do not correlate with the number of salt bridges neutralised. Thus, the charges not included in salt bridges also make a significant contribution to the electrostatic interaction between the two proteins and to the electrostatic complementarity at their interface. This is seen, for example, in the case of the thrombin--hirudin complex, which involves the interaction between five negatively charged residues on hirudin (Asp55, Glu57, Glu58, Glu61 and Glu62) with three positively charged residues at the fibrinogen-recognition exosite of thrombin (Lys432, Arg45 and Arg68). Biologically, this interaction involves the ordering of the carboxy terminus of hirudin upon association (Folkes et al., 1989; Haruyama & WuÈ thrich, 1989; Rydel et al., 1991) and is reflected in a very negative $\Delta G_{\text{Coval}}$ and relatively high electrostatic complementarity. Two of these negatively charged residues and the three positively charged residues also form salt bridges. Removing these salt bridges, and four other salt bridges that are also present in the complex does not dramatically reduce EC or increase $\Delta G_{\text{int}}$ and $\Delta G_{\text{Coval}}$ remains quite negative.

### Electrostatics at the common epitopes of two NA–antibody complexes

The two neuraminidase–antibody complexes 1NMB (NA–NC10) and 1NCA (NA–NC41) have overlapping epitopes on NA. The antibodies NC41 and NC10 bury 899 Å² and 716 Å², respectively, of surface on the neuraminidase, 594 Å² of which is common to both interfaces (Malby et al., 1994). The electrostatic complementarity at the interfaces in these two antigen–antibody complexes is representative of the other protein–protein complexes studied here (see Table 2). Furthermore, this complementarity is also expressed on the subset of the two interfaces that is common to both. Over the common interface the value of EC (Pearson’s correlation for partially solvated proteins) is 0.79 for the NA–NC41 complex and 0.65 for the NA–NC10 complex. We have used the methods described here to investigate similarities in the electrostatic potentials on the common interface surfaces in these two complexes.

The neuraminidase antigens in the two complexes are not identical. In 1NCA, NC41 is complexed with NA from an avian influenza virus (tern N9) and in 1NMB the NC10 is complexed with NA from an influenza virus found in a whale (whale N9). Twelve sequence differences distinguish the two antigens, and whilst none of these are in either the NC41 or the NC10 binding sites, three of the differences involve changes in charge. Furthermore, the three-dimensional structures at the common interface for the two complexes are not identical because of differences in conformation of interface amino acid side chains. Also, the antibodies NC41 and NC10 have very different sequences in their complementarity determining regions. Malby et al. (1994) have shown that the chemical environments of particular antigen residues in the common binding site are quite different in the two complexes. There are two questions to be addressed: how similar are the electrostatic potentials of the two antigen structures on the common binding site, and are there similarities in the potentials of the two quite different antibodies on this common surface?

The two complexes were overlaid by overlapping the Cα positions of the two neuraminidases. The r.m.s. deviation was 0.3 Å. The molecular surfaces

### Table 4. Electrostatic complementarity at selected protein–protein interfaces with salt bridges neutralised

<table>
<thead>
<tr>
<th>PDB</th>
<th>Protein 1</th>
<th>Protein 2</th>
<th>$EC_{\text{int}}$</th>
<th>$EC_{\text{Coval}}$</th>
<th>$\Delta G_{\text{int}}$</th>
<th>$\Delta G_{\text{Coval}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1NCA*</td>
<td>NA</td>
<td>NC41 Fab</td>
<td>0.16 ± 2</td>
<td>0.22 ± 3</td>
<td>0.55 ± 3</td>
<td>0.37 ± 3</td>
</tr>
<tr>
<td>1NMB*</td>
<td>NA</td>
<td>NC10</td>
<td>0.09 ± 1</td>
<td>0.16 ± 5</td>
<td>0.39 ± 4</td>
<td>0.20 ± 7</td>
</tr>
<tr>
<td>2HFL*</td>
<td>HEL</td>
<td>HyHEL5 Fab</td>
<td>0.09 ± 2</td>
<td>0.22 ± 6</td>
<td>0.17 ± 2</td>
<td>0.10 ± 4</td>
</tr>
<tr>
<td>2PTC*</td>
<td>Trypsin</td>
<td>BPTI</td>
<td>0.07 ± 5</td>
<td>0.26 ± 6</td>
<td>0.32 ± 4</td>
<td>0.44 ± 5</td>
</tr>
<tr>
<td>3HFM*</td>
<td>HEL</td>
<td>HyHEL10 Fab</td>
<td>0.08 ± 4</td>
<td>0.12 ± 5</td>
<td>0.27 ± 2</td>
<td>0.34 ± 4</td>
</tr>
<tr>
<td>3HHRf</td>
<td>hGH</td>
<td>hGH₅₉</td>
<td>0.10 ± 2</td>
<td>0.35 ± 6</td>
<td>0.37 ± 1</td>
<td>0.38 ± 1</td>
</tr>
<tr>
<td>3HHRg</td>
<td>hGH/hGH₅₉</td>
<td>hGH₅₉₂</td>
<td>0.22 ± 9</td>
<td>0.25 ± 3</td>
<td>0.58 ± 3</td>
<td>0.47 ± 1</td>
</tr>
<tr>
<td>4HTCb</td>
<td>Hirudin</td>
<td>Thrombin</td>
<td>0.16 ± 2</td>
<td>0.57 ± 2</td>
<td>0.51 ± 2</td>
<td>0.54 ± 2</td>
</tr>
</tbody>
</table>

Residues involved in salt bridges were taken as those in which the distance for the separation of their charged groups was less than 4.0 Å.

Values are given as EC ± Δ where EC+Δ and EC − Δ are correlations r1 and r2, Δ given in the last decimal place. Energies in kcal mol⁻¹ at 25 C.

* de Vos et al. (1992). Salt bridges Asp8–Asp126, and Glu44–Arg16 neutralised.
of whale N9 and tern N9 buried in the interfaces with NC10 and NC41, respectively, were calculated. The overlapped region of these two epitopes were identified as those points on the two surfaces within 1 Å of each other. There were thus two (very similar) surfaces defining the overlapped region, one following the molecular surface of whale N9 and the other following the molecular surface of tern N9. The values of the correlation between these two surfaces were averaged for the calculation of EC, as described earlier.

The first line of Table 5 shows the values of EC when the electrostatic potential on the overlapping epitope is generated by tern N9 and whale N9. The EC values are negative, implying that the electrostatic surfaces are similar. (This can also be seen by comparing the top left and bottom left surfaces of Figure 3.) However, the EC values of greater than −1.0 (exact similarity) show that the movement of side-chain and main-chain atoms, and the net charge difference between tern and whale N9, combine to have a measurable effect on the pattern of electrostatic potential.

The second line of Table 5 shows that the EC values between the potentials generated by the antibodies at the overlapping epitope are insignificant. Thus, although the two complexes display electrostatic complementarity, and the common antigenic sites display electrostatic similarity, there is no measurable similarity in the electrostatic potentials of the two antibodies across the common binding site. The mathematical correlations that exist between the potentials of each antigen–antibody pair must derive from different subsets of the potentials at the common binding site. This becomes evident from comparing the top right and bottom right surfaces in Figure 3, i.e. the dominant contribution to the complementarity in the common epitope interface is due to different spatial regions of the interface.

Table 5. Electrostatic complementarity at overlapping area of epitopes on NA for NC41 and NC10

<table>
<thead>
<tr>
<th>Protein 1</th>
<th>Protein 2</th>
<th>$EC_{T_d}$</th>
<th>$EC_{T_p}$</th>
<th>$EC_{PS}$</th>
<th>$EC_{PS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TN9–1NCA*</td>
<td>WN9–1NMB*</td>
<td>−0.34 ± 1</td>
<td>−0.62 ± 4</td>
<td>−0.55 ± 4</td>
<td>−0.64 ± 1</td>
</tr>
<tr>
<td>NC41–1NCA*</td>
<td>NC10–1NMB*</td>
<td>−0.10 ± 10</td>
<td>−0.10 ± 8</td>
<td>−0.06 ± 6</td>
<td>−0.10 ± 8</td>
</tr>
</tbody>
</table>

* Tern N9 of the 1NCA (tern N9–NC41) complex.
* Whale N9 of the 1NMB (whale N9–NC10) complex.
* NC41 of the 1NCA (tern N9–NC41) complex.
* NC10 of the 1NMB (whale N9–NC10) complex.

Conclusions

Values of $CC^p$ and $CC^c$ are insignificantly small for all the protein–protein complexes studied, and similar results are obtained for a variety of cut-off distances and choice of charge parameters. These results substantially disprove the qualitative assertion often made in the literature that intermolecular interactions involve associations between surfaces with complementary surface charge distributions. Therefore, the term charge complementarity is not appropriate to describe protein–protein interfaces, when used in the sense quantitatively measured by CC.

On the other hand, EC values for all the selected protein–protein complexes are positive. This substantially confirms the qualitative assertion made in the literature that intermolecular interactions involve associations between surfaces with complementary electrostatic potential. The
The term electrostatic complementarity is thus appropriate to describe protein–protein interfaces. All protein charges, and not just the salt bridges or surface charges, contribute towards the value of electrostatic complementarity at the interface. However, charges of atoms more distant from the interface contribute less, so that antigen–Fab and antigen–Fv complexes have similar EC values.

For all the protein–protein complexes studied, $E_{CP}^{PS}$ is greater than $E_{CP}^{FS}$. The greater complementarity of the partially solvated surface electrostatic potentials over the fully solvated ones is also clearly evident from viewing the electrostatic potential (Figure 2). $E_{CP}^{PS}$ also converges more quickly with increasing grid resolution than $E_{CP}^{FS}$. The greater significance of $E_{CP}^{PS}$ values may be because the partially solvated electrostatic potential more accurately represents the bound state of the proteins than the fully solvated electrostatic potential, since the dielectric boundary (although not the charge distribution) is that of the complex. Since $E_{CP}^{PS}$ and $E_{CP}^{FS}$ are correlated, there is little difference in the ordering of the complexes by either measure. Future studies of protein–protein interaction should therefore consider $E_{CP}^{PS}$ and $E_{CP}^{FS}$ to be robust and rigorous measures of electrostatic complementarity.

$E_{C}$ can also be used to measure electrostatic similarity. This principle was applied to compare the electrostatic potential of NA-binding antibodies NC10 and NC41 where their epitopes overlap. The resulting $E_{C}$ value was insignificant, even though NC10 and NC41 both have quite high values of $E_{C}$ for their interaction with NA individually. Thus two surfaces with no measurable similarity in electrostatic potential are able to bind a common target surface. Electrostatic com-

![Figure 3](image-url)
Electrostatic Complementarity

581

plementarity in the two complexes is achieved because it derives from different subsets of the common binding site.

Materials and Methods

Models

Except where stated, all protein atoms and bound calcium ions reported in the crystal structure were included, while crystallographic water molecules and carbohydrate moieties were not. Water molecules and carbohydrate moieties on the surface of the proteins were treated as bulk solvent. Hydrogen atoms were included in the force field. A distance-dependent dielectric of 1.0

were added and the positions determined by energy minimisation over 100 steps of steepest descent using the CVFF (consistent valence force field; Dauber-Osguthorpe et al., 1988) in Discover (Biosym/MSI Technologies Inc.) version 2.95. No Morse or cross terms were included in the force field. A distance-dependent dielectric of 1.0 \(* r\) was used. All aspartate, glutamate, arginine and lysine residues, and both amino and carboxy-terminal groups were considered ionised. Amino and carboxyl-terminal groups, which arose from discontinuity in the co-ordinates where there was disorder in the protein structure, were not considered ionised.

Method for calculation of CC

Atoms within the specified cut-off distance of each other on the proteins forming the complex were identified using the program X-PLOR version 3.1 (Brünger, 1992). Charges were assigned to the atoms in the list using either the PARSE (Skitkoff et al., 1994) or the Formal charge parameter sets. Correlations were then calculated using Pearson’s and Spearman’s rank correlation tests as described below.

Method for calculation of electrostatic energies and EC

The co-ordinates of the points mapping the interface surfaces were determined by MS (Connolly, 1983). DelPhi was then used to generate the electrostatic potential at each point on these buried molecular surfaces. The protein or protein–protein complex was modelled as a low dielectric medium (dielectric constant of 2) and the surrounding solvent as a high dielectric medium, (dielectric constant of 80). The nature of the dielectric variation and the charge distribution were determined by the atomic radii and partial charge parameters assigned to each of the protein atoms. The PARSE charge and radii parameter set (Skitkoff et al., 1994) was used. Ionic strength was zero, as increasing it to physiological strength has been found to have little effect on the resulting solution (Jackson & Sternberg, 1995). The dielectric boundary and charges were mapped onto a cubic grid of size of 201 \(* 201 \times 201\) points/side, with a percentage grid fill of 90%. The electrostatic potential on the boundary of the grid was given by the Debye-Hückel electrostatic potential approximation. Dummy atoms for the uncharged protein were used to maintain the scaling and orientation on the grid of the complex when only one protein was charged. The molecular surface of the proteins was defined using a probe radius of 1.4 Å. The linear Poisson-Boltzmann equation was then solved iteratively until convergence was reached using the QDIFFXS algorithm of version 3.0 of DelPhi (Gilson & Honig, 1988; Nicholls & Honig, 1991). The number of cycles to convergence was automatically determined by the program, and monitored by examining a plot of the convergence in the output file. DelPhi was run on a Convex-C3220, each calculation took up to 1.5 hours to complete.

DelPhi outputs the electrostatic energy \(\Delta G_{\text{Coul}}\) for a protein or protein–protein complex, and enables the calculation of \(\Delta G_{\text{elec}}\) by listing the electrostatic potential generated by the charged, partially desolvated protein at the positions of the charges of the other protein.

Chau & Dean (1994) also used Pearson’s and Spearman’s correlation coefficients to analyse the electrostatic complementarity (in protein-small molecule ligand complexes). However, our method of calculation differs in several significant aspects. The EC value in this work is an average of correlations of the electrostatic potential generated by the proteins on the two interface surfaces while Chau & Dean (1994) only considered the correlation of the electrostatic potential on one surface, that of the ligand. The method for calculating the electrostatic potential of the proteins was also notably different: Chau & Dean (1994) using CNDO/2 charges and constant and distance-dependent dielectrics to calculate the electrostatic potential with the VSS program. PARSE charges are more appropriate for investigation of electrostatic properties than are the Mulliken point charges from very approximate semiempirical molecular orbital theory that Chau & Dean (1994) used. Furthermore, systems where the dielectric constant and the charge density vary from one region of space to another, such as the protein complexes in aqueous solution studied here, are better modelled by the Poisson-Boltzmann equation. It has been shown (for example, see Honig & Nicholls, 1995), that the form of the electrostatic potential is quite dependent on the shape of the dielectric boundary separating the low dielectric used to model the protein from the high dielectric used to model the solvent. Also, in this work we calculate the electrostatic potential surrounding the proteins with and without the dielectric modelling the other protein of the complex present, to give the “partially” and “fully” solvated electrostatic potentials, respectively. Finally, the molecular surface used here was the buried molecular surface as defined by Connolly (1983), rather than an algorithm that took an interface point as any position on the ligand surface that was within 2.8 Å of a site van der Waals surface atom, as used by Chau & Dean (1994).

Dependence of electrostatic complementarity on the parameters of the model

The co-ordinates of the anti-NA antibody NC10 in complex with NA subtype N9 (PDB entry 1NMB; Malby et al., 1994) were used to examine the dependence of EC, \(\Delta G_{\text{elec}}\) and \(\Delta G_{\text{Coul}}\) on the various electrostatic parameters of the model, and the results of this examination are given in Table 3.

The electrostatic energies are not altered by increasing the sampling of the buried molecular surface from 4 grids/Å² to 16 grids/Å². Sampling the interface at the higher resolution does not appreciably change EC, although the time to calculate the electrostatic potential
Electrostatic Complementarity

Statistical analysis

Pearson's correlation coefficient

Pearson’s correlation coefficient ($r^p$) is also known as the linear correlation coefficient, or product-moment correlation coefficient. For pairs of charges or electrostatic potential $a$ and $b$, where $a_i$ and $b_i$ are the charges or electrostatic potential at point $i$, for $i=1 \ldots n$ then:

$$r^p = \frac{\sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{n}(a_i - \bar{a})^2 \sum_{i=1}^{n}(b_i - \bar{b})^2}}$$

where $\bar{a}$ is the mean of $a_i$, $i=1 \ldots n$, and $\bar{b}$ is the mean of $b_i$, $i=1 \ldots n$.

An $r^2$ value of +1 means that the variance of one set of charges or electrostatic potential is perfectly correlated with the variance of the other and $r^2$ value of −1 means that the variances are perfectly anticorrelated. Pearson’s correlations between the charges and electrostatic potential were calculated using the S-Plus package (Statistical Sciences Inc.) version 4.3, running on a Convex-C3220 computer.

Spearman’s rank correlation coefficient

Spearman’s rank correlation coefficient ($r^s$) is a measure of the similarity in pattern between two sets (1 and 2) of charges or electrostatic potential. Each value of the charge or electrostatic potential at the point $i$, $i=1 \ldots n$, is calculated and ranked. Let $R_i$ be the rank of the charge or electrostatic potential at point $i$ among the charges or electrostatic potential in set 1, and $S_i$ be the rank of the charge or electrostatic potential at point $i$ in set 2. These numbers form a uniform distribution of integers between 1 and $n$. Where there are ties in the ranking, the charges or electrostatic potential is assigned the
mean of the rank that they would have had if their values had been slightly different. This rank will either be in integer or a half-integer. The rank correlation is then given by:

\[
r^s = \frac{(n^3 - n - T_1 - T_2) - 6 \sum_{i=1}^{n} (R_i - S_i)^2}{\sqrt{(n^2 - n - T_1)(n^2 - n - T_2)}}
\]

where

\[
T_1 = \sum_{k}(r_k^2 - r_k); \quad T_2 = \sum_{m}(s_m^2 - s_m)
\]

and \(r_k\) is the number of ties in the \(k\)th group of ties among the \(R\) values and \(s_m\) is the number of ties in the \(m\)th group of ties amongst the \(S\) values. The measure is only dependent on the rank of the values (not the actual values), and will be high when the trend is monotonic but not necessarily linear. An \(r^s\) value of +1 indicates correlation and a value of −1 indicates anticorrelation. Spearman's rank correlation was calculated using subroutine G02BQF of the NAG (Numerical Algorithms Group) Subroutine Fortran Library, version 15, run on a Cray Y-MP 4E/464.

Acknowledgements

We thank Michael C. Lawrence and Brian J. Smith for valuable discussions.

References


