

## Impact of a Poisson–Boltzmann electrostatic restraint on protein structures refined at medium resolution

Andrei Korostelev,<sup>a,b</sup> Marcia O. Fenley<sup>b</sup> and Michael S. Chapman<sup>a,b\*</sup><sup>a</sup>Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306-4380, USA, and <sup>b</sup>Kasha Laboratory of Biophysics, Florida State University, Tallahassee, FL 32306-4380, USA

Correspondence e-mail: chapman@sb.fsu.edu

Received 29 April 2004  
Accepted 4 August 2004

The Poisson–Boltzmann formalism has been developed as a restraint for electrostatic interactions during the crystallographic refinement of macromolecules. It accounts implicitly for the effects of solvent and mobile ions, which are usually not included as restraints in the refinement of experimental structures. The electrostatic restraint has been implemented by combining software for numerically solving the three-dimensional Poisson–Boltzmann equation with a package for stereochemically restrained refinement. Its application to medium-resolution protein structures leads to a reduced free *R* factor, overfitting and to improved interactions in salt bridges and between polar or charged amino acids and the solvent. In contrast, Coulombic and screened Coulombic treatments did not lead to significant gains. The work leads to a modest improvement in refinement methods, confirmation that the Poisson–Boltzmann formalism is more consistent with experimental structure than the Coulombic approach, and to a reduction in the discrepancy between experimental and electrostatically optimized atomic models.

## 1. Introduction

Many aspects of biomolecular science depend on the quality of macromolecular structures determined experimentally. The quality of structures, in turn, may depend on the choice of the force fields that are used to restrain stereochemistry during refinement (Hendrickson, 1985; Brünger, 1992*a*). These functions estimate the potential energy with respect to the relative positions of atoms and have terms designed to approximate bonding and non-bonding interactions with surrounding atoms (Weiner & Kollman, 1981; Brooks *et al.*, 1983; Karplus, 1987; Dinur & Hagler, 1991). There are variations in the potential energy functions of different force fields, but the overall stereochemical potential energy function ( $E_{\text{geometry}}$ ) has the following form:

$$E_{\text{geometry}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{elec}} + E_{\text{vdW}} + E_{\text{hb}}, \quad (1)$$

where  $E_{\text{bond}}$ ,  $E_{\text{angle}}$  and  $E_{\text{dihedral}}$  represent covalent bond, angle and dihedral angle energy terms, and  $E_{\text{elec}}$ ,  $E_{\text{vdW}}$  and  $E_{\text{hb}}$  are electrostatic, van der Waals and hydrogen-bond potentials, respectively.

In crystallographic or NMR refinement the stereochemical potential energy is minimized jointly with a target function,  $E_{\text{experiment}}$ , that is minimized when there is good agreement between model and both experimental data and an *a priori* understanding of stereochemical geometry,

$$E_{\text{total}} = E_{\text{geometry}} + w_a E_{\text{experiment}}, \quad (2)$$

where  $w_a$  is a weight chosen to give the best structure as measured by cross-validation (Brünger, 1992a).

There are several force fields that are used in structure determination and molecular simulations. Most have been parameterized against experimental data or *ab initio* calculations of model systems. Valence (or bonding) energy parameters are often derived from small-molecule crystallographic or spectroscopic data or from quantum-mechanics calculations (Lifson & Stern, 1982; Brooks *et al.*, 1983; Nemethy *et al.*, 1983; Hermans *et al.*, 1984; Weiner *et al.*, 1984, 1986; Nilsson & Karplus, 1986; Engh & Huber, 1991). The force fields used in structure refinement have been modified *ad hoc* to improve convergence on the joint stereochemical/experimental target function (Post & Dardarlat, 2001). For example, explicit hydrogen-bonding and electrostatic terms are often omitted. Recently, it was found that an improved hydrogen-bonding term could actually help in the refinement of medium-resolution structures (Fabiola *et al.*, 2002). Here, the potential benefits of improved electrostatic treatments are investigated.

Electrostatic interactions play a vital role in the structural and functional properties of proteins. Accurate simulations require explicit all-atom treatment of solvent molecules. Their large number, uncertainties in their locations and their dynamic distribution lead to high computational complexity. This has motivated continued development of simpler macroscopic (or implicit solvent) approaches, including the generalized Born and Poisson–Boltzmann approximations. Here, the solvent is represented as a featureless dielectric medium and the solute is treated at an all-atom level. Of the implicit solvent methods, the analytical approach of the generalized Born treatment has the advantage of rapid computation (Bashford & Case, 2000). It yields good agreement with experimental  $pK_a$  shifts in small molecules (Luo *et al.*, 1998), binding affinities of small molecules in chloroform (Luo, Head *et al.*, 1999) and ion-pairing in aqueous solution (Luo, David *et al.*, 1999). Calculations with macromolecules are less accurate (Luo *et al.*, 1998; David *et al.*, 2000). The Poisson–Boltzmann approach captures well the electrostatic contributions to the solvation energies of amino acids (Nina *et al.*, 1997) or small solutes (Jean-Charles *et al.*, 1991; Sitkoff *et al.*, 1994) when compared with explicit solvent calculations. The Poisson–Boltzmann treatment has been used to calculate thermodynamic properties of biomolecules, such as  $pK_a$  values and binding free energies (Gilson & Honig, 1987; Penfold *et al.*, 1998; Juffer & Vogel, 2000; Nielsen & McCammon, 2003). These findings suggest that different biophysical properties may be better represented by one or another implicit solvent approximation.

Poisson–Boltzmann calculations have been incorporated into force fields for molecular-dynamics and solvation free-energy studies (Gilson *et al.*, 1995; Luo *et al.*, 2002) and have been coupled to quantum-chemistry methods (Gogonea & Merz, 2000). Such applications have been reviewed by Fogolari, Honig and coworkers (Honig *et al.*, 1993; Fogolari *et al.*, 2002).

The Poisson–Boltzmann equation (PBE) has the form

$$\nabla[\varepsilon(\mathbf{r})\nabla\varphi(\mathbf{r})] = -4\pi\{\rho(\mathbf{r}) + \sum_i q_i n_i \exp[-\beta q_i \varphi(\mathbf{r}) - \nu(\mathbf{r})]\}, \quad (3)$$

where  $\varphi(\mathbf{r})$  is the electrostatic potential at point  $\mathbf{r}$  and  $\varepsilon(\mathbf{r})$  is the position-dependent dielectric constant; on the right side of the equation,  $\rho(\mathbf{r})$  represents the fixed charge density of the solute and the second term accounts for the charge density of the positive and negative mobile ions. These are assumed to be in thermal equilibrium at temperature  $T$  ( $\beta = 1/k_B T$ , where  $k_B$  is the Boltzmann constant);  $n_i$  is the number density of ions of type  $i$  in the bulk solution. The potential energy of charge  $q_i$  located at position  $\mathbf{r}$  is given by  $q_i \varphi(\mathbf{r})$ ;  $\nu$  is a dimensionless excluded volume potential, *i.e.*  $\nu = 0$  in regions of space that are accessible to the mobile ions and  $\nu = \infty$  in regions that are inaccessible to them.

Analytical solutions to the Poisson–Boltzmann equation are only available for cases with idealized geometries such as spheres, cylinders and ellipsoids (Yoon & Kim, 1989; Hsu & Liu, 1996*a,b*). For complex biological molecules with arbitrary charge distributions and surfaces, the Poisson–Boltzmann equation is solved numerically. More efficient and accurate algorithms continue to be developed (Baker *et al.*, 2001; Boschitsch *et al.*, 2002; Boschitsch & Fenley, 2004). One common simplification is to use the first-order term in a Taylor series expansion of the Boltzmann factor, leading to the linearized Poisson–Boltzmann equation:

$$\nabla[\varepsilon(\mathbf{r})\nabla\varphi(\mathbf{r})] = -4\pi\{\rho(\mathbf{r}) + \sum_i q_i^2 n_i \beta \varphi(\mathbf{r}) \exp[-\nu(\mathbf{r})]\}. \quad (4)$$

The linearized form is appropriate in many cases and reduces the computation. However, the linear PBE is a close approximation only when  $\varphi$  is small, which is not true for some biomolecules, including nucleic acids.

Optimization of the parameters used in the Poisson–Boltzmann equation has been attempted through comparison of the calculated  $pK_a$  values of charged residues with experiment (Demchuk & Wade, 1996). Protein interior dielectric is one of the most controversial parameters (Warshel, 1979; Warshel & Russell, 1984; Schutz & Warshel, 2001; Lee *et al.*, 2002). It is generally assumed that the dielectric of proteins is lower than that of bulk water. It should be greater than that of vacuum, but a wide range of relative dielectric constants from 2 to more than 100 have been reported (Gilson & Honig, 1987; Sternberg *et al.*, 1987; King *et al.*, 1991; Pethig, 1992; Smith *et al.*, 1993; Antosiewicz *et al.*, 1994), depending in part on the extent to which atoms, charges, permanent and induced dipoles cannot be explicitly modeled in the system. Values are difficult to assign *a priori* and may depend on the location within a protein (King *et al.*, 1991). Therefore, the spatially averaged values used in practice may need to be optimized for the model parameterizations of individual systems. Other semi-empirical parameters include atomic radii and charges and ways to define the molecular surface (Dong & Zhou, 2002). Many of these parameters are interdependent and adjustable.

In crystallographic refinement, the classical Coulombic potential has sometimes been used as an electrostatic restraint (Brünger, 1992*b*; Brünger *et al.*, 1998). It is also the basis of the

electrostatic terms in molecular-mechanics programs such as *AMBER* and *CHARMM* (Weiner & Kollman, 1981; Brooks *et al.*, 1983). Coulomb's law represents the special case solution to the Poisson equation for two point charges in a uniform dielectric medium,

$$E_{\text{elec}} = q_1 q_2 / (4\pi \epsilon_0 \epsilon r_{12}), \quad (5)$$

where  $r_{12}$  is the distance between charges  $q_1$  and  $q_2$  and  $\epsilon_0$  and  $\epsilon$  are the dielectric constants of vacuum and medium, respectively. Several models were proposed for the effective dielectric constant ( $\epsilon_{\text{eff}}$ ) in proteins to account for solvent screening. Some of them include dependence on the 'exposure' of atoms of interest to the solvent (Mallik *et al.*, 2002). Other dielectric models, such as  $\epsilon_{\text{eff}} = 2 + (r - 1)^2$  (Warshel, 1979; Berman *et al.*, 2000), involve dependence on the distance between the atoms ( $r$ ). Gelin & Karplus (1979) used a dielectric constant that is numerically equal to the interatomic distance in Å ( $\epsilon_{\text{eff}} = r$ ). This simple approach has little theoretical justification. It has been criticized for overestimating the screening of the buried groups and underestimating the screening for the solvent-exposed groups and has shown to be ineffective when scaled with respect to the Poisson–Boltzmann equation (Mallik *et al.*, 2002).

Options for using the Coulombic potential with either constant dielectric or  $\epsilon_{\text{eff}} = \epsilon r$  are offered within some crystallographic refinement packages, such as *X-PLOR* and the *Crystallography and NMR System (CNS)* (Brünger, 1992b; Brünger *et al.*, 1998). However, by default, non-bonded restraints include a repulsive term, but not explicit electrostatics, in *CNS* (Brünger *et al.*, 1998), *SHELX* (Herbst-Irmer & Sheldrick, 1998) and *REFMAC* (Murshudov *et al.*, 1997). The absence of electrostatic terms and attractive van der Waals components reflects a conservative strategy to refinement designed to give the experimental data priority over assumptions implicit in any treatment of stereochemistry. It is also the result of empirical experience that crystallographic refinement works better without the electrostatic restraints currently available. This work started with the suspicion that the detrimental impact of existing restraints may have arisen from poor parameterization of Coulomb's energy or failure to account for other electrostatic components such as the reaction field. However, there are other possibilities. Electrostatic restraints might lock atoms into position prematurely, reducing the convergence radius of refinement. Spurious interactions might arise owing to random error in unrefined coordinates or, more insidiously, owing to the ambiguities in asparagine, glutamine and histidine side-chain conformations, because the positions of O, N and C atoms can often be interchanged without impact upon the fit to the crystallographic data. This latter problem can be mitigated in part with software that resolves the ambiguities by choosing the conformation that maximizes the available hydrogen-bonding interactions (Nielsen *et al.*, 1999; Word *et al.*, 1999; Nielsen & Vriend, 2001).

Omitting electrostatic terms is not very satisfactory, because it is an implicit admission that electrostatics (at a given level of theory) and experimental structure are not fully compatible

and this can cast doubt on detailed structural rationalization of function. Recent implementation of the generalized Born approach as a refinement restraint led to slightly improved refinements and a fuller sampling of conformational space (Moulinier *et al.*, 2003). In our work, the possibility of improving refinement with a more complete Poisson–Boltzmann electrostatic restraint (Luo *et al.*, 1998; David *et al.*, 2000) is tested. Medium-resolution structures are the primary focus since the experimental data alone are insufficient to position the atoms precisely and improved stereochemical restraints have a more measurable direct impact upon model quality.

## 2. Materials and methods

### 2.1. Software design

Electrostatic restraints were calculated with the Poisson–Boltzmann equation. Within this treatment, the total electrostatic force for each atom has four components (Gilson *et al.*, 1995),

$$F_i = F_i^{\text{Coul}} + F_i^{\text{RF}} + F_i^{\text{DBF}} + F_i^{\text{IBF}}, \quad (6)$$

where  $F_i^{\text{Coul}}$  is the Coulombic force computed for all atoms separated from  $i$  by three or more covalent bonds,  $F_i^{\text{RF}}$  is the electrostatic force from the reaction field generated by the solvent environment at atom  $i$ ,  $F_i^{\text{DBF}}$  is the dielectric boundary force acting on atom  $i$  resulting from the tendency of high-dielectric solvent to move into low-dielectric regions and  $F_i^{\text{IBF}}$  describes the ionic boundary force acting on atom  $i$  that is caused by the tendency of mobile ions in solution to move into regions of lower ionic strength.

$F_i^{\text{Coul}}$ , the Coulombic charge–charge interaction force, was calculated using the existing code of the crystallographic refinement package *CNS* (Brünger *et al.*, 1998). A default cutoff of 8.5 Å for intramolecular interactions was applied. The remaining three components of the electrostatic force were calculated with the adaptive Poisson–Boltzmann solver *APBS* 0.2.2 (Baker *et al.*, 2001). With *APBS*, two calculations of the linear Poisson–Boltzmann equation must be performed in order to compute the reaction field force. For the reaction field, the effect of solvent is calculated from the difference between two *APBS* calculations, the first using solute and solvent estimates for  $\epsilon_{\text{in}}$  and  $\epsilon_{\text{out}}$  (respectively) and the second with both set to the dielectric of the solute.

An interface was developed to link *APBS* to the crystallographic refinement program *CNS* as a module (Brünger *et al.*, 1998). The module calculates the electrostatic potential energy and the forces from the derivatives with respect to the position of each atom and this is repeated for each cycle of refinement. Refinement targets and optimization methods of the native *CNS* remain available with the electrostatically restrained version, including least-squares or maximum-likelihood targets, conjugate-gradient or simulated-annealing optimizations, Cartesian or torsion-angle coordinate systems. The default *CNS* non-bonded restraint contains only a repulsive term to account for the van der Waals separation and can

**Table 1**

Medium-resolution structures used in the test refinements.

PDB code	Name	Resolution (Å)	No. amino acids	Reported $R_{\text{free}}$	Reported $R$	Reference
1a43	HIV-1 capsid protein	2.6	72	0.281	0.223	Worthylake <i>et al.</i> (1999)
1awu	Cyclophilin A	2.3	170	0.351	0.316	Vajdos <i>et al.</i> (1997)
N/A†	Snu13	1.9	250	0.461	0.451‡	S. Oruganti, Y. Zhang & H. Li (unpublished work)
1hw6	Apo 2,5-diketo-D-gluconate reductase	2.0	277	0.501	0.487‡	Sanli & Blaber (2001)
2chr	Chloromuconate cycloisomerase	3.0	370	0.264	0.189	Hoier <i>et al.</i> (1994)
1a7b	CD2	3.1	376	0.306	0.239	Murray <i>et al.</i> (1998)
1ab4	Topoisomerase	2.8	477	0.310	0.226	Morais Cabral <i>et al.</i> (1997)
2bct	$\beta$ -Catenin	2.9	502	0.288	0.211	Huber <i>et al.</i> (1997)
1avc	Bovine annexin VI	2.9	648	0.268	0.205	Avila-Sakar <i>et al.</i> (1998)
1a9b	MHC-I-peptide complex	3.2	772	0.305	0.251	Menssen <i>et al.</i> (1999)
6pfk	Phosphofructokinase	2.6	1280	0.255	0.188	Schirmer & Evans (1990)

† Not yet submitted to the PDB. ‡ Unrefined model was used to test restraints during initial refinement.

**Table 2**

High-resolution structures used in the test refinements.

PDB code	Name	Resolution (Å)	No. amino acids	Reported $R_{\text{free}}$	Reported $R$	Reference
1akg	$\alpha$ -Conotoxin	1.1	16	0.157	0.147	Worthylake <i>et al.</i> (1999)
1eb6	Deuterolysine	1.0	177	0.183	0.174	McAuley <i>et al.</i> (2001)
1agy	Cutinase	1.15	197	0.219	0.203	Nicolas <i>et al.</i> (unpublished work)
1ea7	Sphericase	0.93	310	0.193	0.168	Almog <i>et al.</i> (1994)
1m15	Arginine kinase	1.2	361	0.177	0.164	Yousef <i>et al.</i> (2002)
1kdy	AcIPF-proteinase complex	1.1	368	0.182	0.172	Wlodawer <i>et al.</i> (2001)
2gwe	Catalase	0.88	502	0.187	0.184	Murshudov <i>et al.</i> (2002)

therefore be used at the same time with the new electrostatic restraint. The software additions are available at <http://www.sb.fsu.edu/~rsref>.

Parameters for the *APBS* electrostatic calculations were as follows. The solvent-probe radius for molecular-surface determination was 1.4 Å. The van der Waals molecular surface was calculated with harmonic average smoothing. The exterior (solvent) dielectric was 78.4 (for water) and the interior protein dielectric constant was optimized for each protein. The ionic strength was set to zero. The grid spacing depended on the size of the protein and varied from 0.45 to 1.5 Å. The grid was 50% larger than the protein in each dimension in order to avoid edge artifacts.

For analysis of the refined structures, the molecular surface areas of amino acids were calculated with *MSMS* (Sanner *et al.*, 1996). Solvent accessibility was defined as the in-protein accessible area of residue *X* divided by the accessible area in the tripeptide Gly-*X*-Gly in an extended conformation. A residue was treated as surface if its accessibility was more than 50% (Radzicka & Wolfenden, 1988). Salt bridges were counted with a cutoff of 4 Å between a side-chain N atom (arginine or lysine) and an O atom (glutamate or aspartate) (Barlow & Thornton, 1983; Kumar & Nussinov, 2002). Hydrogen bonds were identified with *Hbond2002* (Fabiola *et al.*, 2002) according to distance and acceptor-angle criteria.

## 2.2. Parameterization and test systems

A series of refinements were performed on nine medium-resolution (Table 1) and seven high-resolution protein structures (Table 2) downloaded from the PDB (Berman *et al.*,

2000). The size of proteins varied from 72 to 1200 residues for medium-resolution proteins and from 16 to 500 residues for high-resolution proteins. All non-H protein atoms, water molecules, ions, prosthetic groups and ligands were included. Quasi-Newton optimizers (Liu & Nocedal, 1989) were used within *CNS* (Brünger *et al.*, 1998). For each structure, optimal values were found for  $w_a$ , the weight on the crystallographic data, and  $\epsilon$ , the protein interior dielectric, by two-dimensional grid searches for the best  $R_{\text{free}}$ . Test-set reflections were as defined in the PDB depositions.

## 3. Results and discussion

### 3.1. Refinement with the Coulombic potential

In contrast to the Poisson–Boltzmann treatment (§3.2), a Coulombic restraint, on average, does not improve refinement compared with the default that has no electrostatic restraint, although there is variation among test proteins, with some improving marginally. All refinements were optimized with respect to the dielectric of the Coulombic restraint (Table 3). In four test structures the optimal dielectric constant was 2–3 and the restraint improved refinement marginally. More often, a low dielectric constant leads to a worse free  $R$  factor ( $R_{\text{free}}$ ). Refinement with a dielectric constant larger than 12 is usually indistinguishable from refinement without an electrostatic restraint (Fig. 1). The dielectric constant is effectively a reciprocal weight on the electrostatic restraint relative to other stereochemical restraints.

A Coulombic restraint had the greatest beneficial impact with initial hitherto unrefined models. The initial model of the

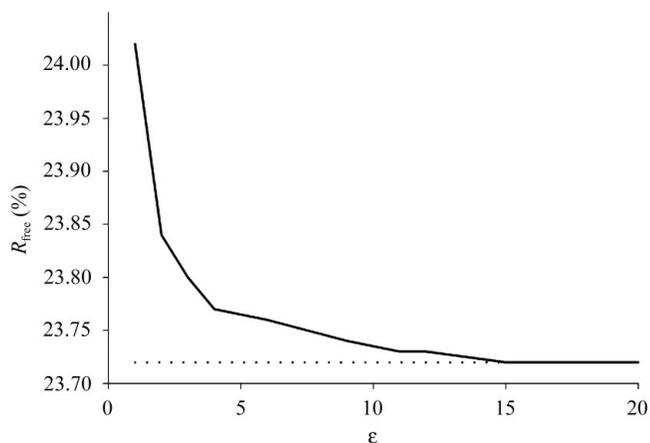
**Table 3**  
Comparison of medium-resolution refinements with or without electrostatic restraints.

PDB code	Refinement without an electrostatic restraint			Refinement with Coulombic restraint†			Refinement with a Poisson–Boltzmann restraint†		
	$R_{\text{free}}$	$R_{\ddagger}$	$R_{\text{free}} - R$	$R_{\text{free}}$	$R$	$R_{\text{free}} - R$	$R_{\text{free}}$	$R$	$R_{\text{free}} - R$
1a43	0.2670	0.2101	0.0569	0.2660	0.2050	0.061	0.2558	0.2280	0.0278
1awu	0.3541	0.3499	0.0042	0.3427	0.3447	−0.002	0.3411	0.3284	0.0127
Snu13§	0.4072	0.3581	0.0491	0.4042	0.3537	0.0505	0.4029	0.3563	0.0466
1hw6	0.4297	0.3750	0.0547	0.4280	0.3721	0.0559	0.4264	0.3725	0.0539
2chr	0.2610	0.1800	0.0810	0.2609	0.1801	0.0808	0.2512	0.1839	0.0673
1a7b	0.3143	0.2287	0.0856	0.3128	0.2298	0.0830	0.3036	0.2480	0.0556
1ab4	0.3122	0.2297	0.0825	0.3108	0.2291	0.0817	0.3109	0.2283	0.0826
2bct	0.2820	0.2100	0.0720	0.2816	0.2095	0.0721	0.2813	0.2100	0.0713
1avc	0.2730	0.1965	0.0765	0.2722	0.1981	0.0741	0.2709	0.1980	0.0729
1a9b	0.3082	0.2573	0.0509	0.3166	0.2532	0.0634	0.3012	0.2573	0.0439
6pfk	0.2372	0.1762	0.0610	0.2372	0.1770	0.0590	0.2372	0.1778	0.0594
Average change with restraint	0.0	—	0.0	−0.001	—	0.0004	−0.006	—	−0.007

‡  $R$  factors were calculated in a consistent manner for all test structures and differ slightly from those reported by the original authors (Table 1), which were calculated with a variety of parameters and scaling procedures. † Coulombic and Poisson–Boltzmann parameters were optimized against each test protein individually. § Not yet submitted to the PDB.

apo-2,5-diketo-D-gluconate reductase refined to  $R/R_{\text{free}}$  of 0.372/0.428 compared with 0.375/0.430 for refinement without the restraint and there was an identical improvement in Snu13 (Table 3). In these tests on unrefined structures, simulated-annealing torsion-angle dynamics was used (Brünger *et al.*, 1999). In all the simulated dynamics refinements, the starting temperature was set to 2000 K and dropped 50 K per cycle.

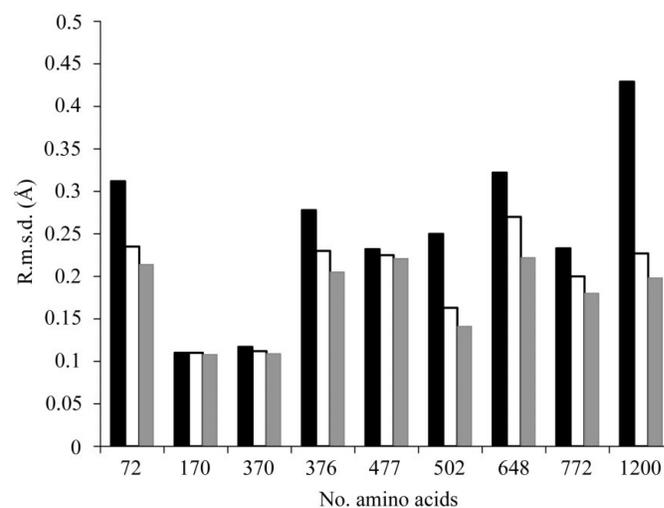
Refinements with a distance-dependent dielectric function,  $\epsilon_{\text{eff}}(\mathbf{r}) = \epsilon \mathbf{r}$ , where  $\mathbf{r}$  is the interatomic distance, produced results similar to those with a constant dielectric (data not shown). Three structures improved slightly: those with low optimized dielectric constants. For the Coulombic treatments with both constant and distance-dependent dielectric constants, it was tractable to incorporate electrostatic interactions between molecules related by crystallographic symmetry. However, this did not lead to significant improvement.



**Figure 1**  
Refinement of phosphofructokinase (PDB code 6pfk) using a Coulombic restraint with varying dielectric constants. Refinement with a large dielectric constant is indistinguishable from refinement without the restraint (dotted line). The refinement is degraded with a lower dielectric, which gives the Coulombic restraint greater weight. This behavior is typical for the Coulombic formalism, explaining why electrostatic refinements have generally not been used in crystallographic refinements.

### 3.2. Refinement with the Poisson–Boltzmann restraint

The restraint was tested in three situations: (i) medium-resolution structures previously refined by conventional methods, (ii) medium-resolution structures as yet unrefined and (iii) high-resolution structures that had been previously refined by conventional methods. Discussion starts with the previously refined medium-resolution structures. These were modestly improved, but to varying extents (Table 3). Free  $R$  factors were lowered by an average of 0.006 and  $\Delta R$ , the difference between  $R_{\text{free}}$  and  $R$ , a measure of overfitting (Brünger, 1992a, 1997), by 0.007. In most cases stereochemistry (root-mean-square deviations of bonds and angles from ideal) was slightly improved. Improvement is inversely dependent on the size of the protein. For proteins under 400 residues,  $R_{\text{free}}$  improved by 0.010 to 0.014 and  $R$  (overfitting)



**Figure 2**  
Structural differences between medium-resolution models refined with and without a Poisson–Boltzmann restraint. All-atom root-mean-square differences (r.m.s.d.s) are shown for three sets of residues: surface-charged (black bars), buried uncharged (grey bars) and all residues (white bars). There is variation in the amount that coordinates are changed with the restraint, but it is consistently surface-charged amino acids that are affected most.

**Table 4**  
Comparison of high-resolution refinements with or without the new electrostatic restraint.

PDB code	Refinement without the electrostatic restraint			Refinement with the optimized Poisson–Boltzmann restraint			
	$R_{\text{free}}$	$R$	$R_{\text{free}} - R$	$R_{\text{free}}$	$R$	$R_{\text{free}} - R$	$\epsilon$
1akg	0.1535	0.1410	0.0125	0.1487	0.1513	-0.0026†	0.1
1eb6	0.1888	0.1744	0.0144	0.1885	0.1743	0.0142	6
1m15	0.1768	0.1637	0.0131	0.1765	0.1636	0.0129	3–4
1agy	0.2149	0.1944	0.0205	0.2146	0.1952	0.0194	3
1ea7	0.1945	0.1621	0.0324	0.1946	0.1621	0.0325	3–5
1kdy	0.1812	0.1674	0.0138	0.1810	0.1669	0.0141	3–5
2gwe	0.1881	0.1814	0.0067	0.1882	0.1814	0.0068	2–4

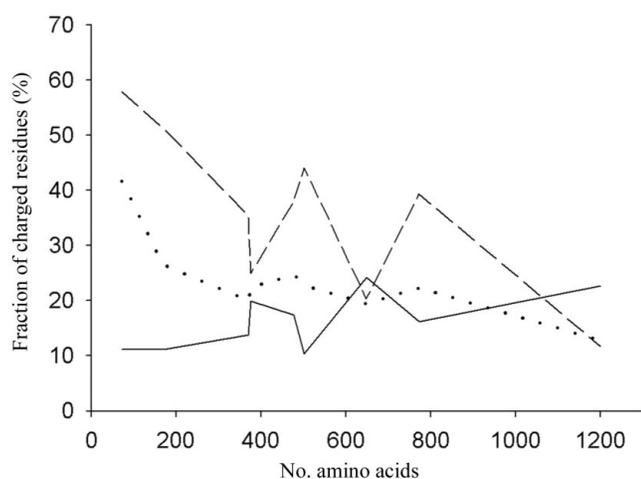
† Small and negative differences between  $R_{\text{free}}$  and  $R$  can arise in several ways. (i) The standard error on  $R_{\text{free}}$  is proportional to  $1/(\text{test-set size})^{1/2}$  (Brünger, 1997). Small proteins may have unavoidably small test sets (351 reflections for 1akg) and hence high random errors for  $R_{\text{free}}$ . (ii) Small test sets can also lead to a systematically low  $R_{\text{free}}$ . Scaling procedures in common use have a significant number of floating parameters (for anisotropy, solvent corrections *etc.*) that can be slightly overfitted and more so against the smaller test set than the working set. This can be avoided with two-parameter Wilson scaling (Fabiola & Chapman, unpublished work), but here it was important to replicate the scaling and statistics of the original published refinements. (iii) Finally, in the high-resolution limit,  $R_{\text{free}}$  is expected to come closer to  $R$  (Brünger, 1997).

by up to 0.03. For larger proteins,  $R_{\text{free}}$  improved by 0.001–0.007 and  $R$  by 0.001–0.008. The size dependence is likely to be because charged surface amino acids are impacted most by the new restraint (Fig. 2) and the proportion of these decreases with the size of the protein (Fig. 3). The size dependence might also suggest a correlation with solvation, but once crystal contacts are accounted for there is not a correlation between the improvement and solvent exposure in the crystal.

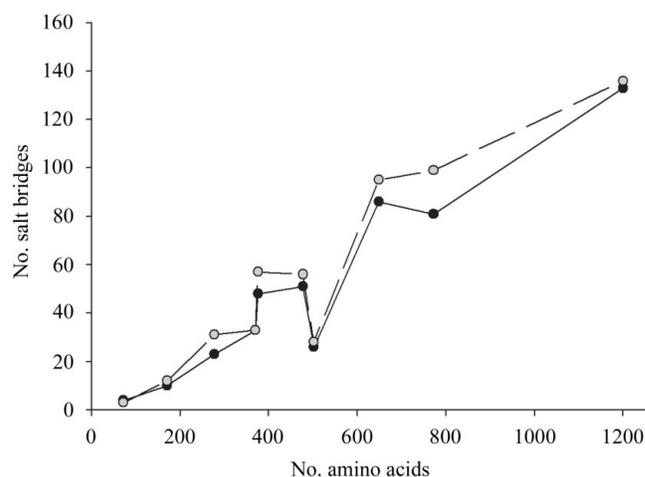
The restraint's impact upon unrefined structures was tested using Snu13 and apo-2,5-diketo-D-gluconate reductase (PDB code 1hw6). With the simulated-annealing refinement that is usually applied during initial refinement, both Coulombic and Poisson–Boltzmann restraints were beneficial, but the Poisson–Boltzmann formulation led to greater improvements (Table 3). The improvement of unrefined structures was slightly less than that of previously refined structures. This might appear counterintuitive, but perhaps reflects the community's earlier experience that early imposition of

(Coulombic) restraints can prematurely lock some conformations that still need to be improved. These results also indicate that the best strategy may be to start refinement without the restraint and to add it as the structure improved.

As expected, for structures already refined at high resolution, the Poisson–Boltzmann restraint had little impact (Table 4).  $R_{\text{free}}$  was reduced by at most 0.005 (for the smallest protein,  $\alpha$ -conotoxin; Fainzilber *et al.*, 1994), while  $\Delta R$  (overfitting) was reduced by 0.015.  $\alpha$ -Conotoxin appears to be an unusual case, because for other refinements at high resolution the impact was more marginal. Less impact is expected than at low resolution, because the additional experimental data available with high-resolution diffraction already defines the atomic positions well. Stereochemical restraints are less important than for medium-resolution structures. Our experience here echoes that with a new hydrogen-bonding restraint (Fabiola *et al.*, 2002). The lack of impact at high resolution (particularly negative impact) shows that the restraints are compatible with accurate structure. Thus, by adding the restraint to medium-resolution refinements, structures are encouraged to adopt features found at high resolution.



**Figure 3**  
Distribution of charged residues in the test proteins: surface (dashed line) and buried (solid line). The proportion of all residues that are on the protein surface is shown as a dotted line. With the finding that surface charged residues are affected most by the new restraint (Fig. 2) and the decreasing number of surface-charged residues with protein size (this figure), there is an explanation for the finding that it is small proteins that are most beneficially impacted by the new electrostatic restraint.



**Figure 4**  
Increased number of salt bridges in medium-resolution refinements with (dashed) and without (solid) the electrostatic Poisson–Boltzmann restraint.

The Poisson–Boltzmann restraint can be compared with the generalized Born restraint developed concurrently by Mouligner *et al.* (2003). Both approaches improve the refinements of medium-resolution structures modestly. In relative terms, the Poisson–Boltzmann approach pays strong dividends, with a threefold greater lowering of  $R_{\text{free}}$ . In absolute terms, the difference in  $R_{\text{free}}$  between the Poisson–Boltzmann and generalized Born approaches (0.006 and 0.002 on average, respectively) may seem small except, perhaps, in the context of the efforts usually made in achieving the last 0.5%  $R$ -factor improvement in a typical structure refinement. The improvement may arise at the cost of greater computational complexity. The results of the two approaches are similar in that the residues most affected are located on the protein surface.

### 3.3. Poisson–Boltzmann parameters

The protein dielectric constant was optimized for each structure by grid search for the lowest  $R_{\text{free}}$ . Optimal values were 1 or 2 for most structures. Optimized values may be lower than the real dielectric because the constant also serves as a reciprocal weight on the electrostatic restraint. Equivalent results could be obtained with higher dielectric constants if an additional explicit weighting parameter was introduced. Increased weighting (lowered dielectric constant) was probably beneficial, because previously refined structures would be in a local minimum of the electrostatically unrestrained target function and would be likely to need perturbation to explore other regions of the energy surface that were electrostatically more favorable.

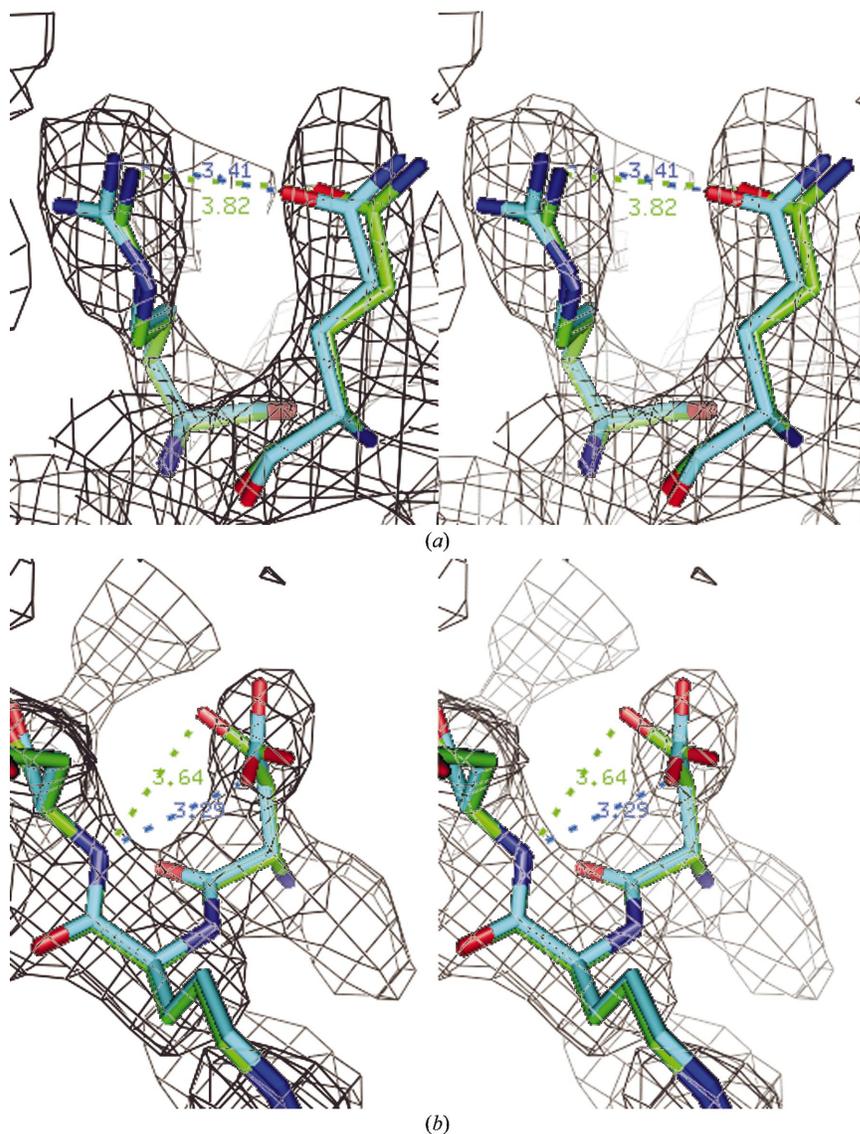
The atomic radii used here are those of the *CNS* and other crystallographic refinement programs (Brünger *et al.*, 1998). They are extended to account for the average van der Waals repulsive effects of unseen H atoms. As the appropriateness of extended radii had not been established for this application, the sensitivity of refinement to these parameters was tested by reducing the radii by 25%. These tests showed that refinements were insensitive to the exact choice of radii.

### 3.4. Computational issues

The adaptive Poisson–Boltzmann solver *APBS* (Baker *et al.*, 2001) offers several boundary-condition options for approximating the effects of electrostatics beyond the limits of the grid. Two were tested: (i) an analytical Debye–Hückel expression for a

single sphere with the molecule’s radius of gyration and net charge and (ii) the sum of Debye–Hückel expressions for spherical ions at each atom. As expected, the greater computation required for the latter was worthwhile, improving  $R_{\text{free}}$  by up to 0.005.

Regarding symmetry-related neighboring molecules, it was easy to incorporate only the Coulombic component of the electrostatics. Reaction-field and dielectric boundary effects would have required extensive modification of the underlying *CNS* code. For medium-resolution structures, the Coulombic



**Figure 5**

Typical joint improvement in electrostatic interactions and fit to electron density. The examples are taken from the Poisson–Boltzmann restrained refinement of the MHC-I–peptide complex. The stereoviews show composite simulated-annealing omit maps (Brünger *et al.*, 1998) calculated from a phasing model (green C atoms) that was refined without any electrostatic restraint, so the density is free from bias from the Poisson–Boltzmann restrained model. Superimposed is the Poisson–Boltzmann restrained refinement (cyan C atoms). (a) The unrestrained distance between the side-chain atoms of ArgA234 and GlnA242 of 3.82 Å decreases to 3.41 Å in the electrostatically restrained refinement. (b) The distance between the backbone N atom of residue B99 and the carboxylate of AspB97 is reduced from 3.64 to 3.29 Å with the restraint.

components from neighboring molecules had no impact, while at high resolution they were slightly deleterious.

Multiple conformers are handled in an incomplete way. The underlying *CNS* retains support for multiple conformations. However, the Poisson–Boltzmann electrostatics use potentials derived from only the first (highest occupancy) conformer. Of tractable approaches, this is the most appropriate. A conformationally averaged or mean-field approach would attenuate the electrostatic effects in a meaningless way. Rigorously, one would have to refine with electrostatics calculated separately for each of the large number of combinations of interdependent selections of local conformers, a feat that does not look tractable in the foreseeable future.

The Poisson–Boltzmann restraint adds a computational overhead to *CNS*. However, *APBS* is fast enough that it is not unreasonable. Without the electrostatic restraint, a batch of gradient-descent refinement (250 cycles) of chloromuconate cycloisomerase (376 residues; Hoier *et al.*, 1994) took 26 min of CPU time (1.8 GHz CPU, 1 GB RAM). The Poisson–Boltzmann restraint calculated on a 0.8 Å grid added 120 min.

### 3.5. Biochemical impact

Use of the Poisson–Boltzmann restraint changes some of the atomic interactions of the refined proteins or at least our recognition of such interactions. It therefore has the potential to alter our functional interpretation of medium-resolution protein structures. The average number of hydrogen bonds per test protein increased by ten and salt bridges between oppositely charged buried side chains by 5.5 (Fig. 4). Charged interactions become more favorable, with a mean reduction in salt-bridge length of 0.13 Å (Fig. 5). The solvation of charged side chains is enhanced, with a slight increase in their solvent exposure (by 0.5% on average). The full impact of the new electrostatic restraint is not yet realised, because the Poisson–Boltzmann calculation treats each molecule as isolated in solution. The distinction between surfaces that are truly solvent-exposed and those involved in packing contacts is currently ignored.

### 3.6. Outlook

In absolute terms, an improvement of  $R_{\text{free}}$  by  $\sim 0.01$  does not appear large. In the context of a much larger number of atomic interactions restrained by covalent stereochemistry, electrostatics is a modest addition to the overall level of restraint that still has an appreciable impact upon overall model quality. The gains from new electrostatic restraints could be combined with similar gains from hydrogen bonding (Fabiola *et al.*, 2002) and perhaps from other improved restraints in the future for a more substantial overall improvement. Furthermore, it should be emphasized that model quality is not just measured by  $R_{\text{free}}$  but how consistent it is with force fields and how seamlessly it can make the transition to molecular-mechanical analyses of function.

Hong Li and colleagues are thanked for providing the structure of Snu13 prior to publication. The work was

supported in part by the National Science Foundation (DBI98-08098 to MSC and CHE 0137961 to MOF) and by the National Institutes of Health (subproject 1 to MSC of P01 GM64676 to T. A. Cross). Software is available at <http://www.sb.fsu.edu/~rsref>.

### References

- Almog, O., Klein, D., Braun, S. & Shoham, G. (1994). *J. Mol. Biol.* **235**, 760–762.
- Antosiewicz, J., McCammon, J. A. & Gilson, M. K. (1994). *J. Mol. Biol.* **238**, 415–436.
- Avila-Sakar, A. J., Creutz, C. E. & Kretsinger, R. H. (1998). *Biochim. Biophys. Acta*, **1387**, 103–116.
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
- Barlow, D. J. & Thornton, J. M. (1983). *J. Mol. Biol.* **168**, 867–885.
- Bashford, D. & Case, D. A. (2000). *Annu. Rev. Phys. Chem.* **51**, 129–152.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
- Boschitsch, A. H. & Fenley, M. O. (2004). *J. Comput. Chem.* **25**, 935–955.
- Boschitsch, A. H., Fenley, M. O. & Zhou, H.-X. (2002). *J. Phys. Chem. B*, **106**, 2741–2754.
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swamithan, S. & Karplus, M. (1983). *J. Comput. Chem.* **4**, 187–217.
- Brünger, A. T. (1992a). *Nature (London)*, **355**, 472–475.
- Brünger, A. T. (1992b). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, Connecticut, USA.
- Brünger, A. T. (1997). *Methods Enzymol.* **277**, 366–396.
- Brünger, A. T., Adams, P. D., Clore, G. M., Gros, P., Gross-Kunstleve, R. W., Jiang, J.-S., Kurzewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst. D* **54**, 905–921.
- Brünger, A. T., Adams, P. D. & Rice, L. M. (1999). *Prog. Biophys. Mol. Biol.* **72**, 135–155.
- David, L., Luo, R. & Gilson, M. K. (2000). *J. Comput. Chem.* **21**, 295–309.
- Demchuk, E. & Wade, R. C. (1996). *J. Phys. Chem.* **100**, 17373–17387.
- Dinur, U. & Hagler, A. T. (1991). *Review of Computational Chemistry*, edited by K. B. Lipkowitz & D. B. Boyd, pp. 99–164. New York: Wiley-VCH.
- Dong, F. & Zhou, H. X. (2002). *Biophys. J.* **83**, 1341–1347.
- Engl, R. A. & Huber, R. (1991). *Acta Cryst. A* **47**, 392–400.
- Fabiola, F., Bertram, R., Korostelev, A. & Chapman, M. S. (2002). *Protein Sci.* **11**, 1415–1423.
- Fainzilber, M., Hasson, A., Oren, R., Burlingame, A. L., Gordon, D., Spira, M. E. & Zlotkin, E. (1994). *Biochemistry*, **33**, 9523–9529.
- Fogolari, F., Brigo, A. & Molinari, H. (2002). *J. Mol. Recognit.* **15**, 377–392.
- Gelin, B. R. & Karplus, M. (1979). *Biochemistry*, **18**, 1256–1268.
- Gilson, M. K. & Honig, B. H. (1987). *Nature (London)*, **330**, 84–86.
- Gilson, M. K., McCammon, J. A. & Madura, J. (1995). *J. Comput. Chem.* **16**, 1081–1095.
- Gogonea, V. & Merz, K. M. Jr (2000). *J. Chem. Phys.* **112**, 3227–3235.
- Hendrickson, W. H. (1985). *Methods Enzymol.* **115**, 252–270.
- Herbst-Irmer, R. & Sheldrick, G. M. (1998). *Acta Cryst. B* **54**, 443–449.
- Hermans, J. Jr, Berendsen, W. F., van Gunsteren, W. F. & Postma, J. P. M. (1984). *Biopolymers*, **23**, 1513–1518.
- Hoier, H., Schlomann, M., Hammer, A., Glusker, J. P., Carell, H. L., Goldman, A., Stezowski, J. J. & Heinemann, U. (1994). *Acta Cryst. D* **50**, 75–84.
- Honig, B., Sharp, K. A. & Yang, A.-S. (1993). *J. Phys. Chem.* **97**, 1101–1109.

- Hsu, J.-P. & Liu, B.-T. (1996a). *J. Colloid Interface Sci.* **178**, 785–788. H.
- Hsu, J.-P. & Liu, B.-T. (1996b). *J. Colloid Interface Sci.* **183**, 214–222.
- Huber, A. H., Nelson, W. J. & Weis, W. I. (1997). *Cell*, **90**, 871–882.
- Jean-Charles, A., Nicholls, A., Sharp, K., Honig, B., Tempczyk, A., Hendrickson, T. F. & Still, W. C. (1991). *J. Am. Chem. Soc.* **113**, 1454–1455.
- Juffer, A. H. & Vogel, H. J. (2000). *Proteins*, **41**, 554–567.
- Karplus, M. (1987). *Protein Engineering*, edited by D. L. Oxender & C. F. Fox, pp. 35–44. New York: Alan R. Liss.
- King, G., Lee, F. S. & Warshel, A. (1991). *J. Chem. Phys.* **95**, 4366–4377.
- Kumar, S. & Nussinov, R. (2002). *Biophys. J.* **83**, 1595–1612.
- Lee, K. K., Fitch, C. A. & Garcia-Moreno, E. B. (2002). *Protein Sci.* **11**, 1004–1016.
- Lifson, S. & Stern, P. (1982). *J. Comput. Phys.* **77**, 4542–4550.
- Liu, D. C. & Nocedal, J. (1989). *Math. Program.* **45**, 503–528.
- Luo, R., David, L. & Gilson, M. K. (2002). *J. Comput. Chem.* **23**, 1244–1253.
- Luo, R., David, L., Hung, H., Devaney, J. & Gilson, M. K. (1999). *J. Phys. Chem. B*, **103**, 727–736.
- Luo, R., Head, M. S., Given, J. A. & Gilson, M. K. (1999). *Biophys. Chem.* **78**, 183–193.
- Luo, R., Head, M. S., Moulton, J. & Gilson, M. K. (1998). *J. Am. Chem. Soc.* **120**, 6138–6146.
- McAuley, K. E., Jia-Xing, Y., Dodson, E. J., Lehmebeck, J., Ostergaard, P. R. & Wilson, K. S. (2001). *Acta Cryst.* **D57**, 1571–1578.
- Mallik, B., Masunov, A. & Lazaridis, T. (2002). *J. Comput. Chem.* **23**, 1090–1099.
- Menssen, R., Orth, P., Ziegler, A. & Saenger, W. (1999). *J. Mol. Biol.* **285**, 645–653.
- Morais Cabral, J. H., Jackson, A. P., Smith, C. V., Shikotra, N., Maxwell, A. & Liddington, R. C. (1997). *Nature (London)*, **388**, 903–906.
- Moulinier, L., Case, D. A. & Simonson, T. (2003). *Acta Cryst.* **D59**, 2094–2103.
- Murray, A. J., Head, J. G., Barker, J. J. & Brady, R. L. (1998). *Nature Struct. Biol.* **5**, 778–782.
- Murshudov, G. N., Grebenko, A. I., Brannigan, J. A., Antson, A. A., Barynin, V. V., Dodson, G. G., Dauter, Z., Wilson, K. S. & Melik-Adamyanyan, W. R. (2002). *Acta Cryst.* **D58**, 1972–1982.
- Murshudov, G., Vagin, A. & Dodson, E. (1997). *Acta Cryst.* **D53**, 240–255.
- Nemethy, G., Pottle, M. S. & Scheraga, H. A. (1983). *J. Phys. Chem.* **87**, 1883–1887.
- Nielsen, J. E., Andersen, K. V., Honig, B., Hooft, R. W., Klebe, G., Vriend, G. & Wade, R. C. (1999). *Protein Eng.* **12**, 657–662.
- Nielsen, J. E. & McCammon, J. A. (2003). *Protein Sci.* **12**, 1894–1901.
- Nielsen, J. E. & Vriend, G. (2001). *Proteins*, **43**, 403–412.
- Nilsson, L. & Karplus, M. (1986). *J. Comput. Chem.* **7**, 591–616.
- Nina, M., Beglov, D. & Roux, B. (1997). *J. Phys. Chem.* **101**, 5239–5248.
- Penfold, R., Warwicker, J. & Jonsson, B. (1998). *J. Phys. Chem. B*, **102**, 8599–8610.
- Pethig, R. (1992). *Annu. Rev. Phys. Chem.* **43**, 177–205.
- Post, C. B. & Dardarlat, V. M. (2001). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, pp. 489–496. Dordrecht: Kluwer Academic Publishers.
- Radzicka, A. & Wolfenden, R. (1988). *Biochemistry*, **27**, 1664–1670.
- Sanli, G. & Blaber, M. (2001). *J. Mol. Biol.* **309**, 1209–1218.
- Sanner, M. F., Olson, A. J. & Spehner, J. C. (1996). *Biopolymers*, **38**, 305–320.
- Schirmer, T. & Evans, P. R. (1990). *Nature (London)*, **343**, 140–145.
- Schutz, C. N. & Warshel, A. (2001). *Proteins*, **33**, 400–417.
- Sitkoff, D., Sharp, K. A. & Honig, B. (1994). *J. Phys. Chem.* **98**, 1978–1988.
- Smith, P. E., Brunne, R. M., Mark, A. E. & van Gunsteren, W. F. (1993). *J. Phys. Chem.* **97**, 2009–2014.
- Sternberg, M. J., Hayes, F. R., Russell, A. J., Thomas, P. G. & Ferscht, A. R. (1987). *Nature (London)*, **330**, 86–88.
- Vajdos, F. F., Yoo, S., Houseweart, M., Sundqvist, W. I. & Hill, C. P. (1997). *Protein Sci.* **6**, 2297–2307.
- Warshel, A. (1979). *Photochem. Photobiol.* **30**, 285–290.
- Warshel, A. & Russell, S. T. (1984). *Q. Rev. Biophys.* **17**, 283–422.
- Weiner, P. K. & Kollman, P. A. (1981). *J. Comput. Chem.* **2**, 287–299.
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. J. & Weiner, P. (1984). *J. Am. Chem. Soc.* **106**, 765–784.
- Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986). *J. Comput. Chem.* **7**, 230–252.
- Wlodawer, A., Li, M., Gustchina, A., Dauter, Z., Uchida, K., Oyama, H., Goldfarb, N. E., Dunn, B. M. & Oda, K. (2001). *Biochemistry*, **40**, 15602–15611.
- Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (1999). *J. Mol. Biol.* **285**, 1735–1747.
- Worthylake, D. K., Wang, H., Yoo, S., Sundqvist, W. I. & Hill, C. P. (1999). *Acta Cryst.* **D55**, 85–92.
- Yoon, B. J. & Kim, S. (1989). *J. Colloid Interface Sci.* **128**, 275–288.
- Yousef, M. S., Fabiola, F., Gattis, J., Somasundaram, T. & Chapman, M. S. (2002). *Acta Cryst.* **D58**, 2009–2017.