calculated structure-factor amplitudes; where

$$wE_{\rm Xray} = w\sum_{hkl}(|F_o| - k|F_c|)^2,$$

where w and k are scale factors.

20.2.4. Empirical parameterization of the force field

Considerable effort has gone into the development of a number of force fields for use in molecular-dynamics simulations of biomolecules (Jorgensen & Tirado-Rives, 1988; MacKerell *et al.*, 1995, 1998; van Gunsteren *et al.*, 1996). The parameters described here are those of the *CHARMM*22 force field, the force field used in the *X-PLOR* program. Estimation of the force constants, equilibrium values and non-bonding parameters in equations (20.2.3.2) and (20.2.3.3) involves a self-consistent approach that balances the bonding and non-bonding interaction terms among the macromolecule and solvent molecules (MacKerell *et al.*, 1998). A wide range of data are taken into account during an interactive process of optimization in order to adequately account for the extensive and correlated nature of the parameters in a consistent fashion. Small-molecule model compounds representative of proteins or nucleic acids are considered in detail, and a hierarchical approach is applied to extend the parameters to larger molecules with minimal adjustment at the points of connection.

The empirical basis of the parameters is broad. Gas-phase geometries and crystal structures are used to determine equilibrium bond lengths, bond angles, and dihedral phase and periodicity. Vibrational spectra, primarily from gas-phase infrared and Raman spectroscopy, are used to fit values for the force constants. Torsion-angle terms are estimated from relative energies of different conformers of model compounds, such as 4-ethylimidazole and ethylbenzene, based on gas-phase data. In cases where no satisfactory experimental data are available, *ab initio* calculations are used to obtain the required energy surfaces. Adjustments are made to describe the energy barriers and positions of saddle points, as well as the minimum-energy structures.

Optimization of the non-bonded parameters includes fitting the van der Waals and electrostatic terms of equation (20.2.3.3), while maintaining a balance among the protein–protein, water–water and protein–water interactions. The parameterization of the *CHARMM*22 force field is based on the water model and water–water interactions of the *TIP*3*P* model (Jorgensen *et al.*, 1983). As such, use of this parameter set with another water model will lead to inconsistencies in the balance of intermolecular interactions. Data from dipole moments, heats and free energies of vaporization, solvation and sublimation, and molecular volumes, as well as *ab initio* calculations of interaction energies and geometries are used to optimize intermolecular interactions. Partial charges of atoms are determined by fitting *ab initio* interaction energies and geometries of small-molecule compounds that model the peptide backbone and amino-acid side chains. Magnitudes and directions of dipole-moment values are also used to optimize partial charges. Experimental gas-phase dipole-moment values are used when available, while *ab initio* calculated values are adopted otherwise. The van der Waals parameters are then refined by comparing results of condensed-phase simulations on pure solvents with heats of vaporization and molecular volumes.

The crystallographic restraint term in the potential-energy function, $E_{\rm rest}$, must also be parameterized to optimize the agreement with the experimental structure-factor amplitudes while simultaneously retaining good geometry and non-bonding interactions. Optimization of $E_{\rm rest} = wE_{\rm Xray}$ involves only the estimation of w. Unlike the parameters in $E_{\rm empir}$, w has no physical basis and is usually chosen to make the force due to $E_{\rm rest}$ balance the total force contributed by all terms in $E_{\rm empir}$. As refinement of the structure progresses, these forces, and hence w, necessarily change since the quality, in terms of geometry and non-bonding interactions, of the structure improves and the crystallographic residual is reduced.

20.2.5. Modifications in the force field for structure determination

Simulated-annealing protocols require modification of the parameters to maintain the correct geometry and local structural integrity of the molecule in order to allow heating to very high, non-physical temperatures for several thousand integration steps. Such modifications are acceptable in the case of structure determination since the primary goal is to define the optimum equilibrium structure in best agreement with the crystallographic or NMR data. Simulations intended to reproduce the fluctuations or dynamic properties of the system must employ carefully defined parameters without such modifications. These modifications include substantial increases in the force constants for bond lengths and angles, *e.g.*, factors of two to ten are used in the parameters specified in the *X-PLOR* file parallhdg.pro. A number of improper torsional terms are added to maintain proper chirality.

The specific terms in $E_{\rm nonb}$ are also modified for the purpose of structure determination. In this methodology, the goal is to converge efficiently to a model that satisfies the experimental data, rather than to obtain an accurate description of the conformational surface, such as estimating fluctuations and equilibrium distributions. Alterations in $E_{\rm nonb}$ include the replacement of the computationally expensive $E_{\rm vdW}$ by a quartic or harmonic repulsive term, which prevents steric conflict among atoms, but ignores dispersive attraction. The electrostatic term, $E_{\rm elec}$, is frequently excluded altogether, since the $1/\mathbf{r}$ dependence of the Coulombic potential allows charge interactions to dominate the interatomic forces far from the global minimum in a fashion that hinders movement toward the global minimum. Exclusion of this important physical property of biological systems is possible, because the crystallographic structure factors contain sufficient information to reflect adequately the imprint of electrostatics on the average structure.

20.2.6. Internal dynamics and average structures

It is most often the goal of the structural biologist to define a single average structure of a macromolecule. The well recognized internal motions arising from thermal fluctuations of a macromolecule may be necessary for function, but, nonetheless, the methods of structure determination generally aim to model a single average structure. Internal motions range from the high frequency, small amplitude motions (*i.e.* those modelled by crystallographic B values) to low frequency, larger amplitude motions of loops and whole domains. Some studies (Kuriyan *et al.*, 1986; Post, 1992) have examined the validity of the assumptions about fast timescale motions made by the methods of structure determination. It is reasonable that some of the differences between the structure solutions of a protein obtained by NMR spectroscopy and X-ray crystallography are due to differences in the effects of internal motions. The application of molecular-dynamics algorithms for structure determination has allowed the use of protocols that account for effects of internal motions by employing time-averaged restraints (Schiffer *et al.*, 1995).

references