multimodal distribution was observed for the CE(1,2)—CZ—OH angles, with maxima at 118 and 122° (Fig. 18.3.2.2). The scatter plot of CE1—CZ—OH versus CE2—CZ—OH demonstrates that this distribution typifies individual fragments and does not arise from differing classes of fragments. This justifies an asymmetric parameterization for these angles; symmetric parameterization would require correspondingly soft force constants. The major difference between the histidine parameters listed here compared to those of EH arise from the appearance of HISD (uncharged; unprotonated at NE2) fragments in the CSD. The EH parameterization assumed values from other fragments. The total of 12 fragments is not large, but does predict some alterations in parameters involving the ring nitrogens. The fragment selection reported here did not investigate effects of noncovalent binding. For the aromatic residues, these include hydrogen-bonding effects (especially for histidine) and π-cloud interactions. Appropriate fragments exist in the database, so such dependencies are, in principle, accessible to investigation.

18.3.2.3.3. Aliphatic residues: leucine, isoleucine, valine

Compared to EH parameterization, the only notable features of the aliphatic residues were the leucine bonds and the C—CA—CB angles of isoleucine and valine. The leucine CD—CG(1,2) bonds retained relatively large σ values, which rather increased compared to the previous values. The C—CA—CB angle values, clustered as bare carbon/tetrahedral CH extended atom/tetrahedral CH2 extended atom in EH, are sensitive to the degree of substitution at the CB carbon (Table 18.3.2.3, see the discussion of peptide fragments above). The statistics here show that the EH (1991) parameters were too small by about 2°.

18.3.2.3.4. Neutral polar residues: serine, threonine, glutamine, asparagine

These residues share neutral polarity, but are all geometrically distinct. Like leucine, valine and isoleucine described above, threonine is branched at CB, and the parameterization for C—CA—CB should be chosen accordingly. Additionally for threonine, the CA—CB—CG2 angle, clustered with valine as CH1E—CH1E—CH3E in EH (1991), should be altered from 110.5 to 112.4° according to the statistics reported here. The tabulated glutamine and asparagine parameters are taken from identical amide-group statistics, and parameters for the aliphatic atoms of glutamine are taken from arginine. This choice of fragments arose from a desire to maximize the number of fragments for the amide group; however, the individual residues might be expected to exhibit residue-specific amide structures.

18.3.2.3.5. Acidic residues: glutamate, aspartate

The fragment definitions were chosen to select both symmetrically and asymmetrically encoded carboxylate structures; that is, the statistics include carboxylate groups with delocalized charges as well as carboxylate groups encoded with a single charged oxygen atom. This distribution presumably reflects the variations in proteins as well. For both glutamic and aspartic acids, statistical variation in the asymmetry of delocalization was evident. One measure of parameter variation as a function of varying charge delocalization is the anticorrelation of C—O bond lengths and CH2—C—O bond angles. For example, while the standard deviation of the corresponding aspartate bond lengths individually is 0.024 Å, the standard deviation of their pairwise average is 0.012 Å. Similarly, the standard deviation of the glutamate CH2—C—O bond angles individually is 2.1° but the standard deviation of the pairwise average is 0.6°. This coupling of parameters is an example of additional information potentially available for structure refinement, but which would require new formulations of restraints.

18.3.2.3.6. Basic residues: arginine, lysine

The 98 arginine fragments in the database did not show alterations from the EH values, except generally tighter constraints at the guanidinium group. Lysine CD—CE bond lengths are somewhat shorter in the new statistics, while the two angles derived from the fragments remained similar.

18.3.2.3.7. Sulfur-containing residues: methionine, cysteine, disulfides

One of the most conspicuous features of the EH parameters is the soft force constant for the methionine SD—CE bond length. The 49 fragments now in the CSD also show a sample deviation for the 1.774 Å average bond length of 0.056 Å, and after one 4σ outlier rejection, the tabulated value of 1.779 Å still has a large sample deviation of 0.041 Å. In practice, the use of soft restraints during refinement often leads to warnings of relatively large deviations from the target value. Inspection of the CSD structures did not reveal an artificial source of this greater variability. Cysteines and disulfides here show reduced sample σ values for generally similar average target values.

18.3.2.4. Planarity restraints

Planarity and improper dihedral restraints, being ‘hard’ restraints, are amenable to the same kind of parameterization described above. Physically realistic deviations should be allowed. A survey of several planar atoms, such as CG of aromatic residues, the inter-ring carbons (CD2, CE2) of tryptophan and CZ of tyrosine, showed standard deviations about strict planarity of 1°–2°. Statistically significant deviations of average values from perfect planarity might also be expected, particularly as a function of the protein fold environment. For example, an average nonzero planarity of the ω angle of the peptide bond has