most cases, uncertainty in rejection criteria affected the average values little, but could significantly alter standard deviations.

18.3.2.3. *Bonds and angles*

18.3.2.3.1. *Peptide parameters: proline, glycine, alanine and CB substitution*

Fragments representing five-atom lengths of the backbone currently provide adequate statistics for peptide compositions of varieties including glycine, proline and side chains branched at CB. Peptide cyclicity was generally allowed on the assumption that this does not introduce distortions greater than typical protein secondary-structure interactions. The results are presented in Table 18.3.2.3. With one exception, none of the values deviates from those of 1991 by more than one sample standard deviation. However, the very large σ values for the proline C—N—CA and C—N—CD angles (Table 18.3.2.1) are conspicuous. Using high-resolution protein structures, Lamzin *et al.* (1995) identified geometries of proline that were inconsistent with high-resolution protein structures and also noted inconsistencies in C—CA—CB angle parameters (see also the sections on individual amino acids below). In the case of proline, a bimodal distribution of these parameters could be resolved with the discrimination between *cis* and *trans* forms (Fig. 18.3.2.1). A scatter plot of the angles against ω torsion angle resolves the averages (and σ's) of 122.6 (50) and 125.4 (44)° for C—N—CA and C—N—CD, respectively, into *cis*- and *trans*-dependent values with much smaller sample deviations (see Table 18.3.2.2). The large σ value for CB—CG remains, however, particularly for *trans*-proline. Its origin is unknown, but proline pucker may play a role.

Glycine, with its unique CH_2 as CA, required new atom-type definitions for Engh & Huber (EH) (1991) parameterization to account for parameter-average differences of about one-half of a sample standard deviation. These also included C—N—CA, for which the average angles were 120.6° for glycine and 121.7° for the rest. The new statistics with 83 C—CO—NH—CH_2—C fragments estimate a larger value of 122.3° for the glycine C—N—CA angle.

'Extended atom'-type parameterizations, which cluster carbon atoms according to the number of bound hydrogen atoms, naturally separate parameters involving CB into values representing alanine and branched and unbranched side chains. Separate analyses of the bonds and angles for fragments depending on the number of hydrogen atoms at CB (1, 2 or 3) revealed significant variation for the C—CA—CB and N—CA—CB angles. The fragments chosen for peptide parameterization did not cover all possibilities for the peptide chain. In particular, effects of charges at the termini were not analysed. Also, specific residue sequences likely to have statistical effects, such as Pro-Pro (Bansal & Ananthanarayanan, 1988), were not analysed here. With 50–60 relevant fragments from the predominantly α-helical ROP protein, Vlassi *et al.* (1998) were able to compile statistics for main-chain bonds and angles and compare them with protein refinement parameters. Differences from EH were particularly significant for CO and CA—C bonds (1.237 and 1.508 Å, respectively) and for the O—C—N angle (121.35°). Excepting the proline O—C—N angle, for which the new CSD statistics predict an average value lowered to 121.1°, these values remained relatively unchanged. A likely source of the difference might be the predominantly helical structure of the ROP protein; the helical hydrogen bonding directly involves the C—O group in a systematic way.

18.3.2.3.2. *Aromatic residues: tryptophan, phenylalanine, tyrosine, histidine*

With the exception of generally lower σ values, tryptophan parameters remain essentially unchanged. Phenylalanine, also with generally lower σ values, is also essentially unchanged with the assumption of Gaussian distributions. However, a scatter plot of the CB—CG—CD1 *versus* CB—CG—CD2 angles shows an inverse correlation between these two angles, corresponding to ring rotations about an axis perpendicular to the ring face. Non-Gaussian distributions were most evident for tyrosine. In addition to the phenomenon described for phenylalanine, a clearly 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 CH_2 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 CH_2—C—O bond angles. For example, while the standard deviation

references