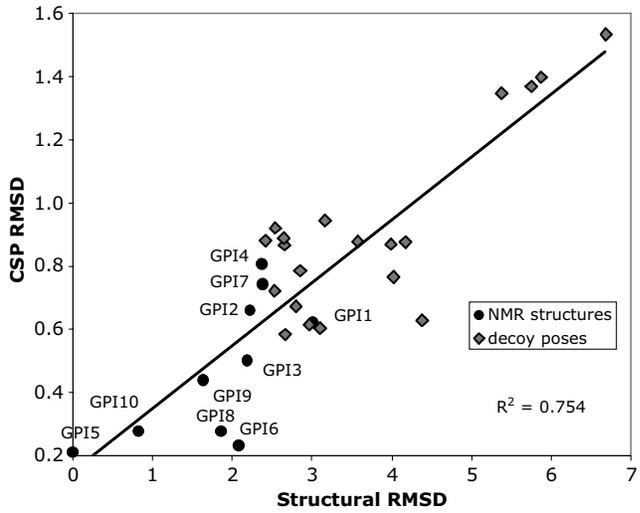

Figure 8.5. Correlation plot of the chemical shift perturbation root-mean-square deviations (rmsds) (in ppm) versus the structural rmsds (in Å) with respect to NMR structure GPI5 for the nine remaining GPI NMR structures and twenty computationally generated structures.

reparameterized the Karplus relationships for Ile, Val, and Thr residues based on DFT calculations to gain improved insights into side-chain dynamics. Trans hydrogen bond scalar J couplings have been detected in nucleic acids and proteins recently. DFT studies of these J coupling constants on peptide models hinted at the cooperative nature of hydrogen bond interactions.[44]

NMR spectroscopy is an important tool to study protein and protein/ligand complexes and novel QM approaches will continue to affect our understanding and interpretation of experimentally observed NMR parameters.

USING QUANTUM MECHANICS TO MODEL PROTEIN STRUCTURE

A major unsolved problem in biology is determining protein structure from sequence (the protein folding problem).[45–47] This includes, in the case of SBDD, predicting the fold from structure or refining homology models. Currently, classical or simplified potentials are used in attempting to solve this problem and significant amounts of very creative effort has been reported, which is beyond the scope of this chapter.[45–47] Instead we focus on the use of QM to discriminate decoy protein structures from native using semiempirical QM methods. Indeed, with the introduction of linear-scaling QM techniques, the modeling of full protein systems is now possible; we briefly review large-scale validations of the method to study protein structure and folding.

Protein geometry validation

To first assess the suitability of semiempirical methods it is important to first demonstrate that the experimental geometries of proteins are reproduced by semiempirical QM theory. Semiempirical methods were developed to handle a large variety of chemical systems by being parameterized against a wide range of small molecules.[48,49] Proteins are large biopolymers with a small number of unique functional groups, so any error in the semiempirical treatment of those particular groups may potentially be magnified in these systems. Although semiempirical methods are heavily parameterized, they differ from classical approaches such as MM approaches that may employ amino-acid-specific parameters.[13]

A quote from Stewart's development of the semiempirical parameterized model 3 (PM3) method captures the essence of this problem nicely, "The parameter set here has three limitations: in the limit, it is only as good as the reference data used; . . . and it should be used with caution when applied to the prediction of any properties not used either in the parameterization or in subsequent surveys."[49] An analysis of the training compounds used in the parameterization of PM3 highlights the absence of several functional groups present in amino acids. Although the amino acids alanine and glycine were included in the parameterization of PM3, guanidyl-like groups and imidazole-like groups were not included.

Large-scale optimization of protein geometries at the semiempirical level highlights the problems of minimizing in vacuo. Unlike classical methods, QM methods can undergo conformational changes as well as changes in bonding configuration. Bond cleavage and proton transfer involving charged groups were common when optimizing with semiempirical methods in vacuo. These artifacts, not suprisingly, were corrected for when optimizing with an implicit solvent model.

Overall, semiempirical methods match the geometries of proteins surprising well. The largest observed discrepancies were in the torsional angles, in agreement with previous observations.[49,50] Furthermore, optimization with SE methods led to a smaller fraction of side chains in native rotameric states. This is likely to be due to the low energetic differences arising from perturbations in the torsional angles. Also of note, the C–N peptide bonds in proteins are longer by 0.06Å in semiempirical-minimized geometries, and in general C–N bonds were predicted to be longer than those found in crystal structures of proteins. Nonetheless, the optimized structures, although far from perfect, reproduced experimental x-ray geometries satisfactorily.

Approximations to semiempirical geometries

Quantum mechanical calculations are more sensitive to geometries of structures than classical methods. Very large changes in the energy can result from small changes in the structure, particularly with respect to bond lengths and angles. In this regard, it is desirable to first optimize a system before applying QM calculations on the structure. Ideally, it would be preferable to optimize the structure at the QM level so that the resulting structure would be consistent