Computational Biology and Chemistry 54 (2015) 33–43

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Tri-peptide reference structures for the calculation of relative solvent accessible surface area in protein amino acid residues

Christopher M. Topham [a,b,c,d,*], Jeremy C. Smith [b,c,d]

[a] Molecular Forces Consulting, 40 Rue Boyssonne, Toulouse 31400, France
[b] Computational Molecular Biophysics, IWR der Universität Heidelberg, Im Neuenheimer Feld 368, Heidelberg D-69120, Germany
[c] University of Tennessee/Oak Ridge National Laboratory, Center for Molecular Biophysics, P.O. Box 2008, Oak Ridge, TN 37831-6309, USA
[d] Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, M407 Walters Life Sciences, 1414 Cumberland Avenue, Knoxville, TN 37996, USA

ARTICLE INFO

Article history:
Received 22 June 2014
Received in revised form 23 November 2014
Accepted 30 November 2014
Available online 3 December 2014

Keywords:
Peptide reference structure
Relative amino acid solvent-accessible surface area
Peptide geometry
Protein conformation
Quantum chemical modelling

ABSTRACT

Relative amino acid residue solvent accessibility values allow the quantitative comparison of atomic solvent-accessible surface areas in different residue types and physical environments in proteins and in protein structural alignments. Geometry-optimised tri-peptide structures in extended solvent-exposed reference conformations have been obtained for 43 amino acid residue types at a high level of quantum chemical theory. Significant increases in side-chain solvent accessibility, offset by reductions in main-chain atom solvent exposure, were observed for standard residue types in partially geometry-optimised structures when compared to non-minimised models built from identical sets of proper dihedral angles abstracted from the literature. Optimisation of proper dihedral angles led most notably to marked increases of up to 54% in proline main-chain atom solvent accessibility compared to literature values. Similar effects were observed for fully-optimised tri-peptides in implicit solvent. The relief of internal strain energy was associated with systematic variation in N, C^{α} and C^{β} atom solvent accessibility across all standard residue types. The results underline the importance of optimisation of 'hard' degrees of freedom (bond lengths and valence bond angles) and improper dihedral angle values from force field or other context-independent reference values, and impact on the use of standardised fixed internal co-ordinate geometry in sampling approaches to the determination of absolute values of protein amino acid residue solvent accessibility. Quantum chemical methods provide a useful and accurate alternative to molecular mechanics methods to perform energy minimisation of peptides containing non-standard (chemically modified) amino acid residues frequently present in experimental protein structure data sets, for which force field parameters may not be available. Reference tri-peptide atomic co-ordinate sets including hydrogen atoms are made freely available.

1. Introduction

Solvent-accessible surface area is a widely used physical environmental parameter in the comparison, prediction and design of protein structure (Hubbard and Blundell, 1987; Šali and Blundell, 1990; Rost and Sander, 1994; Eyal et al., 2004; Worth et al., 2009; Li et al., 2013). Fast surface area calculation methods allow implicit solvent models to be used to take account of solvation energy contributions in scoring protein conformations, or in simulations of protein folding and dynamics (Chen et al., 2008; Vorobjev, 2011; Kleinjung and Fraternali, 2014).

Relative amino acid residue solvent accessibility values provide a convenient means with which to compare solvent accessible surface areas in different residue types, or residues in different protein environments. These values can be calculated by normalising summed atomic solvent-accessible surface areas with respect to summed areas in the same residue type in a common solvent-exposed conformation available to all residue types. The relative amino acid residue solvent accessibility metric affords a basis upon which different structural regions in proteins can be defined (Levy, 2010). It has been found that evolutionary rates of residues at protein interfaces depend on the extent of surface burial (Eames and Kortemme, 2009; Franzosa and Xia, 2009), and relative residue accessibility is an important parameter in the prediction of protein interaction sites (Porollo and Meller, 2007; Zellner et al., 2012). Normalisation permits the meaningful comparison of solvent accessibility between different amino acid

* Corresponding author at: Molecular Forces Consulting, 40 rue Boyssonne, Toulouse 31400, France.
E-mail address: christopher.topham@gmail.com (C.M. Topham).

http://dx.doi.org/10.1016/j.compbiolchem.2014.11.007
1476-9271/

residue types, for example, at structurally aligned positions within protein families and superfamilies (Mizuguchi et al., 1998; Holm and Rosenström, 2010). Relative accessibility-based physical environmental class descriptors are used in protein structural alignment algorithms (Madhusudhan et al., 2009; Topham et al., 2013) and as conditional feature variables in multi-dimensional amino acid substitution tables with diverse predictive applications (Overington et al., 1992; Topham et al., 1993; Worth et al., 2011).

Tri-peptide models in extended conformations of sequence Ala-X-Ala or Gly-X-Gly where X is the residue of interest, often serve as reference molecules to represent the maximally solvent-exposed state (Lee and Richards, 1971; Chothia, 1976; Satow et al., 1980; Miller et al., 1987; Hubbard and Thornton, 1993; Samanta et al., 2002; Ahmad et al., 2003; Nguyen and Rajapakse, 2005). Extended conformation peptidic fragments containing all the naturally occurring residue types are frequently observed in proteins, and are represented in structural alphabet block descriptions of local protein structure (Etchebest et al., 2005). Molecular dynamics simulations of short terminally-blocked alanine- and glycine-containing peptides carried out in explicit solvent with additional provision for solute polarisation reveal heavy sampling of extended conformations in the β-sheet and polyproline II regions of the Ramachandran plot (Hu et al., 2003; Wang et al., 2006). Recently, Gly-X-Gly tri-peptide reference models in residue-specific conformations not available to all residue-types have been proposed in order to accommodate a minority of surface residue positions in protein structures that yield relative solvent accessibilities in excess of 100% with respect to maximal solvent accessibilities in extended conformations (Singh and Ahmad, 2009; Tien et al., 2013).

Ideally, model tri-peptide structures in the designated reference conformation should exist in or close to (local) minima on potential energy surfaces, or at population maxima identified in statistical analyses of protein conformation databases. Reference tri-peptide atomic co-ordinate sets are generally built using selected values for the main-chain and side-chain proper dihedral angles observed in experimentally determined protein or peptide structures, whilst bond length, valence bond angle and improper dihedral internal co-ordinates are set to molecular mechanics force field or other reference values. However, force field reference values are not true equilibrium values in a minimum energy configuration, but correspond to values adopted when all other terms in the force field are set to zero. Thus unless the full form of the potential energy function is applied to minimise the energy, the accumulation of small deviations from equilibrium values may exert significant effects on inter-atomic separation, and hence solvent accessibility. An analogous argument applies concerning the use of context-independent average bond length and valence bond angle parameter values obtained from analyses of experimental structure databases in which distinct component conformation-dependent data distributions can remain hidden (Berkholz et al., 2009). The fixing of main-chain and side-chain proper dihedral angles to inappropriate or mutually incompatible combinations may also lead to poor estimates of maximal atomic solvent accessibility in reference state models, even when all other degrees of freedom have been optimised.

Given the widespread availability of software packages for geometry optimisation based on either classical molecular mechanics force field or quantum chemical methods, it is perhaps surprising that neither constrained nor full energy minimisation protocols have apparently hitherto been implemented. In consequence, the literature abounds with calculated values of solvent accessible surface areas of amino acid residues in model-built tri-peptides in extended and other conformations. However, reference sets of the atomic co-ordinates themselves are generally not readily available. Not only does this render assessment of tri-peptide structural quality difficult, it also precludes the calculation of surface area using alternative parameter sets, such as solvent probe size and solute van der Waals radii, or the choice of alternative surface types and algorithms other than those used to report the original values. For example, the AGBNP (Gallicchio and Levy, 2004) Generalised Born rapid surface area calculation method is based on a Gaussian description of molecular shape (Grant and Pickup, 1995), and depends upon the use of a smaller increase in the solute van der Waals radii (0.5 Å) than is typically employed (1.4 Å) to represent the radius of a water molecule. In contrast, a larger water probe radius of 2.0 Å used to approximate the dielectric boundary in Poisson–Boltzmann solvation energy calculations was found to give optimal agreement with molecular dynamics results for alanine decapeptide in explicit solvent (Aguilar et al., 2010). Moreover, solvation free energy calculation methods, whether physical–chemical or empirical (e.g. Wang et al., 2001; Hou et al., 2005) in origin, require the presence of both hydrogen and heavy atoms.

Here we report on the application of density functional theory (DFT) quantum chemical modelling methods (Burke, 2012) in the geometry optimisation of solvent-exposed reference-state tri-peptides in extended conformations. These methods take explicit account of electronic structure, and their improved accuracy over pair-wise additive molecular mechanics methods with fixed partial charges that neglect polarisation effects has been demonstrated in benchmarking studies of relative performance on di-peptide (Kaminský and Jensen, 2007) and tetra-peptide (Jiang et al., 2010) systems containing cognate amino acids. Density functionals in popular use do not in general require further parametrisation for main-group chemistry applications. This affords an additional advantage in the treatment of non-standard (chemically modified) amino acid residue types, present in significant number in recombinant protein atomic co-ordinate data sets and structural bioinformatics databases (Chandonia et al., 2002; Garavelli, 2004), and for which molecular mechanics force field parameters are often lacking.

2. Methods

2.1. Quantum chemical calculations

Density functional theory (DFT) allows for a more accurate representation of electron correlation effects than does Hartree–Fock theory at relatively modest increased computational cost. Although DFT methods are less accurate than *ab initio* wave function theory methods, which in principle can provide exact molecular geometries and energies, the use of full electron correlation and completely flexible basis sets is impracticable for tri-peptide systems. The B3LYP density functional used here is one of the first standard exchange-correlation density approximations, but continues to remain the most popular for a broad range of general applications in chemistry (Burke, 2012). Known deficiencies of DFT methods in the handling of dispersion forces, the attractive part of the van der Waals interaction, have more recently led to the development of the Minnesota functionals by Zhao and Truhler (2011). These functionals use the same basic generalised gradient approximations as standard approximations such as B3LYP, and achieve improved performance on energy training sets through the fitting of additional adjustable parameters. Benchmark testing by Jiang et al. (2010) of the B3LYP and Minnesota M05-2X density functionals on five different constrained tetra-peptide geometries yielded similar reference molecular geometries and conformation-dependent energy difference profiles.

The B3LYP hybrid Hartree–Fock/density functional method in the GAUSSIAN 94 program suite (Frisch et al., 1995) was used