doi:10.1016/j.jmb.2005.12.012 J. Mol. Biol. (2006) 356, 928–944

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Conformational Diversity of Ligands Bound to Proteins

Gareth R. Stockwell[1,2] and Janet M. Thornton[1]*

[1]*European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton Cambridgeshire CB10 1SD UK*

[2]*Department of Biochemistry and Molecular Biology University College London Gower Street, London WC1E 6BT, UK*

**Corresponding author*

The phenomenon of molecular recognition, which underpins almost all biological processes, is dynamic, complex and subtle. Establishing an interaction between a pair of molecules involves mutual structural rearrangements guided by a highly convoluted energy landscape, the accurate mapping of which continues to elude us. Increased understanding of the degree to which the conformational space of a ligand is restricted upon binding may have important implications for docking studies, structure refinement and for function prediction methods based on geometrical comparisons of ligands or their binding sites. Here, we present an analysis of the conformational variability exhibited by three of the most ubiquitous biological ligands in nature, ATP, NAD and FAD. First, we demonstrate qualitatively that these ligands bind to proteins in widely varying conformations, including several cases in which parts of the molecule assume energetically unfavourable orientations. Next, by comparing the distribution of bound ligand shapes with the set of all possible molecular conformations, we provide a quantitative assessment of previous observations that ligands tend to unfold when binding to proteins. We show that, while extended forms of ligands are indeed common in ligand–protein structures, instances of ligands in almost maximally compact arrangements can also be found. Thirdly, we compare the conformational variation in two sets of ligand molecules, those bound to homologous proteins, and those bound to unrelated proteins. Although most superfamilies bind ligands in a fairly conserved manner, we find several cases in which significant variation in ligand configuration is observed.

Keywords: ligand; conformational diversity; molecular recognition; superfamily; protein

Introduction

Although small in comparison to proteins, many biological ligands are capable of considerable conformational variability. It is clear that the combinatorial effect of varying just a few rotable bonds leads to a large number of possibilities: if we assume that any one rotable bond can exist in three distinct rotational states, then a molecule with only ten such bonds can theoretically adopt $3^{10}=59{,}049$ different conformations. Of course, many of these will be physically impossible due to steric clashes between atoms. Still others may be strongly disfavoured energetically, but a large number may have an energy not too distant from the global minimum, meaning that a considerable region of conformational space is potentially available for exploration by small organic molecules.

Upon binding to a protein, the conformational freedom of a ligand is typically restricted to a small locale, which is usually distinct from the optimum conformation of the solvated molecule, and in many cases may not even be close to a local energy minimum.[1,2] In some cases however, ligands retain considerable mobility even when complexed with proteins.[3] It must be remembered that while molecular recognition is thought to be driven primarily by enthalpic change, and therefore characterised by the formation of specific interactions between the protein and the immobilised ligand, the favourable entropic effect of maintaining some ligand flexibility can in some cases

Abbreviations used: CSD, Cambridge Structural Database; GR, glutathione reductase; FR, ferredoxin reductase; PDB, Protein Data Bank; NAD nicotinamide adenine dinucleotide; FAD, flavin adenine dinucleotide; DT, diptheria toxin.

E-mail address of the corresponding author: thornton@ebi.ac.uk

compensate for weaker interactions between the two molecules.

Ligand conformation has been analysed previously from a number of different perspectives. In one of the earliest large-scale analyses of ligand conformation, Moodie & Thornton[4] compared the coordinates of protein-bound nucleotides to those in an unbound dataset obtained from the Cambridge Structural Database (CSD).[5] The main finding was that although most rotable bonds did not change significantly between bound and unbound states, larger differences were observed for a minority of bonds. These rotations tended to result in the nucleotides changing from folded to extended conformations upon binding to proteins. The authors suggest that the relatively small number of torsional changes which occur upon ligand binding are evidence that bound ligands exist predominantly in low-energy conformations, since similar conformations in solution have been shown both theoretically and experimentally to be in low-energy states.[6]

This assumption does not hold for ligands in general, however, according to a recent study which assessed the energetic effects of ligand reorganisation upon binding.[7] The energy calculations performed here claim[1,2] that ligands rarely bind in their lowest-energy conformation. Surprisingly, however, this study also found that in only about one-third of cases were the bound conformations within 0.5 kcal mol^{-1} of a "local" energy minimum. Moreover, no correlation was found between global strain energy of the bound ligands and either the number of polar interactions, which they made with their proteins (contradicting the study of Nicklaus *et al.*[1]), nor the binding affinity. These results suggest that even quite energetically costly ligand rearrangements can be tolerated without penalising binding affinity. The authors also report that the ligands with the highest strain energies tended to be those that unfolded to the greatest degree upon binding. This finding, taken together with the lack of correlation between strain energy and number of hydrogen bonds, indicates that the primary stabilising factor offered to ligands by their binding sites is that of a generally hydrophobic environment which protects the uncovered non-polar ligand regions from exposure to solvent.

A comparison of the binding sites for the redox cofactors NAD and NADP showed that, although they are structurally similar, these two molecules exhibit some noticeable differences in their interactions with proteins.[8] The authors report a total of 13 different conformations adopted by the two cofactors: eight of them by NADP and five by NAD. Interestingly, the two different compounds were never found in the same conformational cluster, indicating that even an apparently small chemical elaboration can significantly alter the constraints acting on the shape of a particular molecule. The relationship between the conformational clustering, and evolutionary similarities between the proteins is mentioned only in passing; the authors note that proteins with similar folds and/or functions tend to bind the cofactor in a similar conformation, but do not investigate this in detail.

An analysis of the sequence and structure of 32 families of FAD-binding proteins has been reported.[9] The authors comment, in passing, on the conformation of the cofactors bound to each protein. They find that, while in some structural families (namely the *p*-cresol methylhydroxylase and pyruvate oxidase proteins), the cofactor conformation is essentially fixed, other groups (the glutathione reductase (GR) and ferredoxin reductase (FR) families) contain proteins which bind FAD in quite varied arrangements. This study does not compare the conformational variance of FAD "within" families to that observed "between" families.

Hansen and co-workers published a study of the relationship between protein sequence similarity and NAD cofactor conformation.[10] Comparing clusters of sequences of NAD(P)-utilising enzymes from the SWISS-PROT database[11] with the conformations of bound cofactors, the main result reported is that each sequence family binds NAD(P) molecules in conformations which cluster together. This is not surprising, since the identification of protein relationships from sequence alone implies that more than 30% of their residues are identical, which in turn implies that function is conserved, and hence the cofactor should be expected to bind in the same way. Although relationships between higher level protein similarities (i.e. superfamily or fold groupings) and cofactor conformation are not discussed at length, the overall finding is that members of the large structural families (Rossmann fold oxidoreductases and flavin-NAD(P)-coupled enzymes) both bind their cofactors in several different conformations, while the smaller families each map to just a single conformational cluster.

Increased understanding of the conformational changes which ligands undergo upon binding to proteins is attractive for both academic and pragmatic reasons. Appreciation of the energetic constraints acting on bound ligands can shed light on how they perform their particular biological functions, be these as enzymatic cofactors, signalling molecules or labile substrates. In particular, strain induced in a ligand as a consequence of the shape which it is forced by the protein binding it to adopt may promote its participation in chemical reactions. Alternatively, conformational change upon binding may expose a reactive atom or functional group which would otherwise be inaccessible to other reactants.

From a practical standpoint, the investigation of bound ligand conformations may prove valuable in the development of docking methods, where libraries of rotamers derived from bound ligand structures could be used to speed up conformational searches by *a priori* elimination of some of the large number of possible conformations mentioned previously. Furthermore, knowledge of