

Full Length Research Article

Analysis of structurally conserved atomic interactions in structural homologs of nicotinamide adenine dinucleotide binding dehydrogenases

P. S. Solanki¹, M. Krishna Mohan¹, P. Ghosh¹ and S. L. Kothari²

¹Birla Institute of Scientific Research, Jaipur, India.

²Department of Botany, University of Rajasthan, Jaipur, India.

Accepted 12 March, 2012

The stability of protein-ligand complex is decided by the strength of the forces and interaction that holds the overall structure. The binding of ligand to its receptor involves the participation of atomic interactions of key conserved amino acid residues which are essential for functional and structural integrity of protein molecule. Understanding the binding mechanism of ligand to its receptors is important in exploring the structure activity relationships of any protein. In the present study, a dataset of nicotinamide adenine dinucleotide (NAD) binding dehydrogenase was prepared by screening 19 structural homologues using the SCOP database to study the structurally conserved atomic interactions. An interaction profile of NAD against all structural homologues was determined using protein interaction analyzer and structure parser (PIASP) program and the structurally conserved atomic interactions were identified at 50% cut off level. We identified that out of 44 ligand atoms, only single interaction (ASP: O2B-OD1) was 100% conserved throughout the family members. Thus, it is clear from this study that ligand implements varying degree of orientation for binding with receptor molecule. However, there are certain critical amino acid residues that remain conserved throughout the family and participated in all orientations. However, the interaction of aspartate protease (ASP) residue is critical to the active site of selected protein family. This residue, which is common in active site of all the chosen structures, is found to be structurally conserved throughout the family.

Key words: Structurally conserved atomic interaction, conserved residues, protein interaction analyzer and structure parser (PIASP).

INTRODUCTION

Knowing the structures of biological macromolecules is always useful for predicting, interpreting, modifying, and designing their function. The aim of the structural genomics project is to deliver structural information about most proteins. It is not feasible to determine the structure of all proteins by experiment; useful models can be obtained by fold assignment and comparative modeling of those protein sequences that are related to at least one known protein structure. Since each enzyme is specific in their action, the active site of enzyme is

structurally a very important part of the proteins, as it determines the binding proficiency of a protein. In all the members of a given protein family, the residues which forms the active site should remain conserved, however, not necessarily that all the residues should participate in binding. It was observed that few of them play a crucial role in binding with the ligand (Kasinos et al., 1992).

The Protein Data Bank (PDB) is the most important database with structural information of biological macromolecules, with over 77,000 protein structures. The majority of structures in PDB are resolved either through X-ray crystallography or NMR imaging techniques (Frances et al., 1977; Helen et al., 2000, 2004). X-ray crystallography enables details of covalent and non

*Corresponding author. E-mail: pooransingh@gmail.com.

covalent interactions to be analyzed quantitatively in three dimensions, thus providing the basis for the understanding of binding of ligands to proteins (Palmer and Niwa, 2003). The PDB file of any biological complex contains diversified information in the form of record types exhibiting various attributes of protein structure such as atomic coordinates, protein sequence, secondary structural information, experimental remarks. Based upon the structural data, several tools have been developed to analyze protein-ligand interactions such as PLID (Reddy et al., 2008), LigBase (Stuart et al., 2002), 3DinSight (Jianghong et al., 1998), SitesBase (Nicola et al., 2006), ConSurf (Haim et al., 2010), PDB-Ligand (Jae-Min, 2005), ProLINT (Kitajima et al., 2002), LIGPLOT program (Wallace et al., 1995), and CKAAPs DB (Wilfred et al., 2002), but none of them provides the information on structurally conserved atomic interactions at family level.

The formation of biological complex is fundamental to biological process and functionality of enzymes. The major manipulation may disrupt the functionality of molecules. But to understand the mechanism of molecular recognition between the receptors and ligand, manipulation of complexes is essential for studies on protein engineering and has many applications such as engineered enzymes, biosensors, genetic circuit, signal transduction pathways and chiral separations (Loren et al., 2003). The receptor-ligand interaction is very crucial in determining the functionality of given molecular complex, thus it has always attracted the attention of researchers. It is well known that the functional specificity of protein molecules is due to the structural conservation of three-dimensional structure and the amino acid residues present on surface which participates in catalytic activity. The conformational stability of receptor-ligand complex is governed by the strength of the forces and interactions that holds the overall structure. Among these interactions, H-bonding interaction acting upon the interacting atoms such as Donor-Donor and Donor-Receptor plays a key role in maintaining the structural integrity (Branden and Tooze, 1991).

Structure of NAD binding dehydrogenases

Dehydrogenase family is one of the most studied protein family, and has always been a good model system for carrying out research in structural bioinformatics. Primary sequence comparison has indicated that there is an evolutionary relation between many dehydrogenases. The first ever isolated alcohol dehydrogenase (ADH) was purified in 1937 from *Saccharomyces cerevisiae* (Negelein and Wulff, 1937). The dehydrogenases catalyze the oxidation of alcohol to the carbonyl compound or aldehydes. They utilize the coenzyme called nicotinamide adenine dinucleotide (NAD) for their action and this coenzyme participates in the dehydrogenation reaction (Zhi-Jie et al., 1997). The X-ray crystallographic

structure of dehydrogenases shows that it consists of two domains, namely the nucleotide-binding domain and the catalytic domain, the first domain plays a role in binding the coenzyme, often NAD, and the second domain plays a role in binding the substrate. Since the catalytic domain involves in catalysis it is variable in nature and depends upon the substrate specificity and contains amino acid involving in catalysis (Benyajati et al., 1981). Little sequence similarity has been found in the coenzyme-binding domain although there is a large degree of structural similarity, and it has therefore been suggested that structure of dehydrogenase has arisen through gene fusion of ancestral coenzyme nucleotide sequence with various substrate specific domain. The nucleotide-binding domain is formed from the similar overall folding of the polypeptide chain for all the dehydrogenases. The detailed geometry of this domain varies considerably from one enzyme to another. The nucleotide-binding domain composed of six strands of parallel beta sheet with parallel helix running anti-parallel to the sheet. The dehydrogenases family composed of different members based on their substrate specificity but all the members have similar nucleotide binding domain. For the present investigation only those proteins that possess NAD⁺ binding specificity have been chosen.

Binding site analysis

Characterization of a protein function and understanding the specific nature of a protein's binding as a critical part of both protein engineering and structure based drug discovery. Binding site analysis combines several tools that enable you to identify and characterize a protein-binding site and then use those characteristics to the similar features in other proteins of known structures. Active sites are usually identified through homology with another protein or from biochemical data, but identification by this method is not always possible. Binding site analysis identifies functionally important residues by sequence variation across a family. This information can then help you modify or design drug that target such residues. The numerous successes of structural biology have shown that three-dimensional structure of the proteins adds vital information and insight into understanding a protein function are often found by mapping the result of protein sequence analysis onto a known protein structure with known function. The amino acids in the active site are near in space but not necessarily sequentially close to one another. The active site finding tools in binding site analysis allows us to search for crevices that are large enough to bind a ligand.

The importance of hydrogen bonding

The importance of H-bonding for the structure and function

Table 1. Represents dataset of dehydrogenase structural family (based upon SCOP database).

Enzyme name	No of structures	PDB-ID
Fatty oxidation complex alpha subunit	4	1WDK, 1WDL, 1WDM, 2D3T
GDP-mannose 6-dehydrogenase	2	1MUU, 1MV8
L-3-hydroxyacyl CoA dehydrogenase	8	1F0Y, 1TL0, 1LSJ, 1LSO, 1M75, 1M76, 3HAD, 3HDH
Mannitol 2-dehydrogenase	2	1LJ8, 1M2W
Prephenate dehydrogenase TyrA	2	2G5C, 2PV7
UDP glucose dehydrogenase	1	1DLI

of biomacromolecules has been demonstrated by extensive statistical, experimental and theoretical studies (Legon and Millen, 1992; Buckingham et al., 1988; Baker and Hubbard, 1984; Ippolito et al., 1990; McDonald and Thornton, 1994; Subbarao and Haneef, 1991). The detailed attribution of binding free energy has demonstrated the intrinsic importance of interaction; H-bonding to the detailed mechanisms of binding specificity, stabilization of antibody-antigen and protease-inhibitor complexes in solution (Fersht et al., 1985; Serrano et al., 1993). Although the resolution of the X-ray data is probably not sufficient to establish the complete interfacial H-bond system of protein complexes with absolute certainty (Morris et al., 1992), a significance number of interfacial hydrogen bonds with good distance and angular geometry for interaction, can reasonably be assumed to be formed in the crystal structures of the protease-inhibitor and antibody-antigen complexes within the PDB forming 8 to 13 and an average of ten hydrogen bonds between the docking surfaces. Moreover, only a few of the donor/acceptor atoms involved in these intermolecular interactions are capable of forming intramolecular hydrogen bonds suggesting their principal role is one of functional recognition. By comparison however, the subunit interfaces of oligomeric proteins form far fewer hydrogen bonds in proportion to the docking surface areas, some examples forming none. The majority of interfacial hydrogen bonds between oligomeric subunits involve charged donor or acceptor groups, which should make a significant contribution to the subunit interactions (Janin et al., 1988). The energy of Hydrogen bonding interaction depends critically on the donor-acceptor spatial separation and line of approach of the donor hydrogen to the acceptor lone pairs. Other factors such as specific environment of individual H-bonds and specific arrangement of an interfacial hydrogen bond system have also been demonstrated to be major determinants in the overall hydrogen bond stabilization of protein complexes (Smith-Gill et al., 1982; Fersht et al., 1985). On the basis of the physical nature of the H-bond, the interacting donor and acceptor sites of a protein complex intermolecular hydrogen bonding system are required to have a certain degree of spatial and directional complementarities. A graph theoretic method has been reported for the search of small ligand-protein hydrogen bonding sites based upon graph clique finding

(Subbarao and Haneef, 1991; Smellie et al., 1991). However, the computational time required for clique finding between two large proteins would be prohibitively long. A minimum of two potential intermolecular hydrogen bonds in the complex is required for the prediction of relative orientations of the components.

The present work focuses on analysis of binding sites of NAD in dehydrogenases for identifying the structurally conserved H-bonding interactions. This will help us to find out whether the active site is conserved or not, if conserved then what are the critical residues which are important in binding of NAD with protein. We can also find the superposability of the H-bond interaction in any given pair of binding sites in two different proteins. It will also help us to analyze whether the NAD binds in the same orientation in all the active sites or it binds differently in specific set of protein.

MATERIALS AND METHODS

There are large numbers of PDB structures available in the Brookhaven protein data bank that utilizes NAD as a coenzyme. Since dehydrogenases binding to NAD are most commonly studied complexes, it will be interesting to analyze the active site of NAD binding regions for their structural conservation. NAD being the common ligand in these complexes and its H-bonding interactions can be calculated and analyze structurally conserved interactions.

Dataset preparation

In order to understand the behavior of interaction specificity in protein-ligand complexes of a family, NAD binding dehydrogenases family from SCOP database (Andreeva et al., 2004; Murzin et al., 1995) was selected. The family was 6-phosphogluconate dehydrogenase-like, N-terminal domain [51868]. The family attributes are: beta-sheet is extended to 8 strands, order 32145678; strands 7 and 8 are anti parallel to the rest. The six enzymes (Table 1) from the family carrying NAD as a cofactor were selected and all structural homologs were downloaded from Protein Data Bank. Finally, the dataset comprises 19 structural members of NAD binding dehydrogenases Figure 1.

Definition of hydrogen bond donors and acceptors

The following definition of H-bond donors and acceptors has been used for different atom types. The hydrogen bonding donor and acceptor groups at physiological pH are given in Table 2.

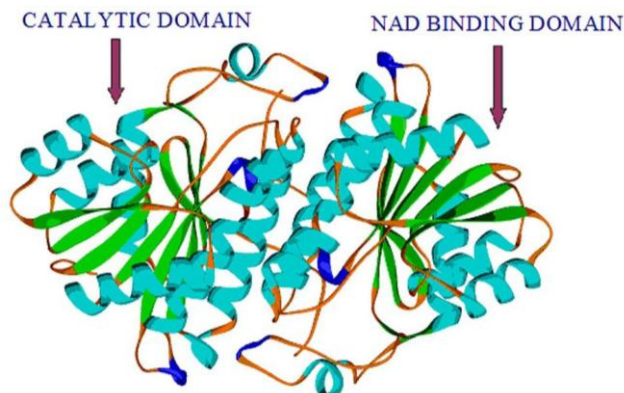


Figure 1. Structure representing catalytic and NAD binding domain of dehydrogenases family.

Table 2. Hydrogen bond donors and acceptors (Torshin et al., 2002; Subbarao and Haneef, 1991).

Donors
(1) N (main-chain N-H)
(2) Asn OD1, His NF2, His ND1, Lys NZ, Asn ND2, Gln NE2, Arg NH1, Arg NH2, Ser OG, Thr OG, Tyr OH, Trp NE1.
Acceptors
(1) O (main-chain C = O)
(2) Asp OD1, Asp OD2, Glu OE2, Glu OE1, His ND1, Asn OD1, Gln NE2, Asn ND2, Ser OG, Thr OG1, Tyr OH

Note: The names are corresponding to PDB id code.

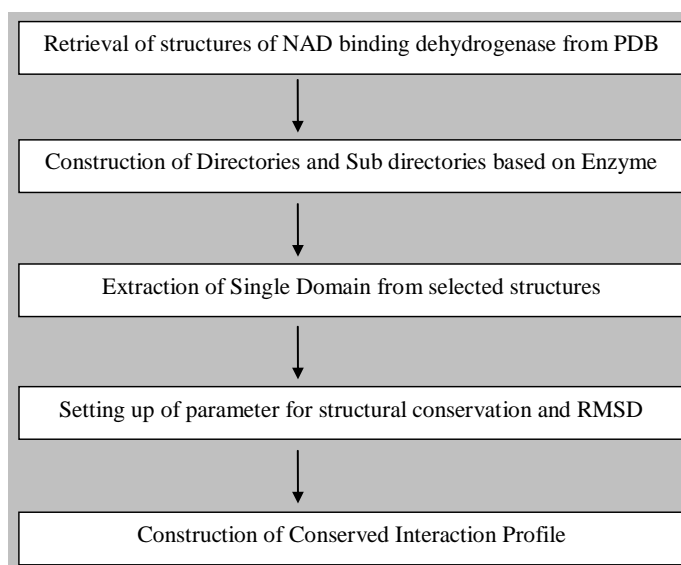


Figure 2. Flowchart of methodology of PIASP.

Analysis of interaction specificity

The analysis of protein-ligand complexes of NAD binding dehydrogenases was performed using in-house developed program, PIASP, stands for Protein Interaction Analyzer and

Structure Parser. PIASP is a web based tool for mining protein-ligand interactions at atomic level. Flowchart of methodology implemented in PIASP is shown in Figure 2. It provides the complete interaction profile of all ligand atoms to its receptor and identifies interaction specificity in all structural homologs. It reports

Table 3. Structurally conserved atomic interactions. Cut-off for structural conservation is 50% and for atomic distance is 3.5 Angstrom. The row in red color indicates the 100% structural conserved of atomic interaction.

Number	Ligand atom	Residue name	Residue atom	Percentage (%)
1	O2A	GLY	CA	63.16
2	O4D	ASN	ND2	52.63
3	O3B	ASP	OD2	89.47
4	O2D	GLU	CD	63.16
5	O2D	GLU	OE2	63.16
6	O5B	GLY	CA	68.42
7	O3D	LYS	NZ	52.63
8	O2B	ASP	OD1	100.00
9	O2B	ASP	CG	57.89

the all conserved interactions essential for Receptor ligand interaction. In addition, PIASP has equipped with various utilities which are very useful for any researcher working on protein-ligand interactions. This tool enhances our understanding of protein ligand interaction at atomic interaction level and could be a great help when designing compounds for molecular docking like drugs. PIASP can be accessed from BISR website (<http://bistr.res.in/cgi-bin/project/piasp/index.cgi>).

All the selected structures were downloaded from PDB and categorized into name based classes for PIASP analysis. The single domain extracted using PIASP was further confirmed using Discovery Studio software from Accelrys Incorporation. The default parameters of structural conservation (that is, 70%) and RMSD (3.5 Angstrom) were chosen for analysis. Based upon the interaction profile of all NAD atoms, it constructs the result table for Structurally Conserved Atomic Interactions (Table 3).

RESULTS AND DISCUSSION

For decades, lot of research work has been done on catalytic mechanism of enzymes and on their binding strategies. The protein-ligand interaction is very crucial in determining the functionality of given protein. The interaction of ligand to its receptors molecule involves the participation of few atoms and the binding strategy is very important in understanding the mechanism of enzyme action. The present work focuses on nicotinamide adenine dinucleotide (NAD), which is one of the most commonly used organic cofactor in living cells, essential for many metabolic processes. Many NAD binding proteins which are responsible for variety of activities in cellular environment have been discovered. The binding of NAD to its receptor proteins is very specific and crucial for its activity. Thus, how NAD binds to various kinds of receptor protein is important to understanding the mechanism of enzyme action.

We investigated the interaction specificity of ligand, NAD, to the active site of its receptors (NAD binding dehydrogenases) using the in-house developed PIASP program. The structural homologues of NAD binding family of dehydrogenases were selected from SCOP database and all respective structure files were downloaded from the Protein Data Bank. The interaction

profile of NAD against all the structural members was built at cutoff distance of 3.5Angstrom and structural Conservation of 50% (Table 3). The interaction profile at family level was generated successfully which represents the interaction of each of the ligand atom against atom of amino acid residues at the set parameter.

The interaction profile shows the strategy adopted by 44 atoms of NAD for interacting with the various atoms of receptor molecules. This profile shows the interaction of particular ligand atom with residual atom based upon the cutoff distance, so it becomes very clear how many atoms of ligands are participating in the binding. It was found that specific atom type of ligand behaves in different manner in different structures.

Based upon the interaction profile, a table of structurally conserved atomic interaction was generated in PIASP. We calculated structural conservation at 50% level (Table 3). It is obvious from the table that out of 44 atoms of NAD there were very few atoms found to participate in interaction with the atoms of amino acid residues of receptor molecule. It is clear from the table that interaction of O2B atom of NAD is 100% structurally conserved with OD1 (ASP) atom in all structural members of the family. Besides, there were few more interactions of NAD atoms (that is, O3B and O5B) which were found to be conserved at 50% level. From the structurally conserved table (Table 3), it may be inferred that the binding of NAD atoms was not exactly identical in all members of the family. NAD binds in different orientation in different structure. No two structures have shown the identical pattern of atomic interaction. Thus, this analysis shows that the binding of ligand to its receptor is very specific and depends upon the geometry of the active site pocket of receptor.

Conclusion

In the present research work, NAD binding protein families have been studied to understand the behavior of interaction specificity of NAD ligand against the structural homologs of dehydrogenases family. It is clear from the

result obtained that different proteins have different binding specificity but there are few atoms of NAD, which participate in binding in all the structures. From interaction profile of NAD, it was clear that out of 44 atoms; only one atomic interaction was 100% structurally conserved throughout the family. The amino acid residue (ASP) is critical to the active site of selected protein family. This residue, which is common in active site of all the chosen structures, is found to be structurally conserved throughout the family. In active sites of all the family, it will present at specific position and also structurally conserved not only in the confined active sites but also in the entire molecular structure, when they are aligned pair wise.

Thus, it can be concluded that ligand implements varying degree of orientation for binding with receptor molecule. However, there were certain critical amino acid residues that remain conserved throughout the family and participated in all orientations.

ACKNOWLEDGEMENT

We are thankful to BTIS Centre funded by Department of Biotechnology, Government of India for providing infrastructure facilities at Birla Institute of Scientific Research, Jaipur.

REFERENCES

- Andreeva A, Howorth D, Brenner SE, Hubbard TJP, Chothia C, Murzin AG (2004). SCOP database in 2004: refinements integrate structure and sequence family data, *Nucleic Acids Research*, 32:226-229.
- Benyajati C, Place AR, Powers DA, Sofer W (1981). Alcohol dehydrogenase gene of *Drosophila melanogaster*: relationship of intervening sequences to functional domains in the protein. *Proc. Natl. Acad. Sci. U.S.A.*, 78(5): 2717–2721.
- Branden C, Tooze J (1991). *Introduction to Protein Structure*. Garland Publishing inc, New York and London.
- Fersht AR, Shi JP, Knill-Jones J, Lowe DM, Wilkinson AJ, Blow DM, Brick P, Carter P, Waye MM, Winter G (1985). Hydrogen bonding and biological specificity analysed by protein engineering. *Nature*, 314(6008):235-8.
- Haim A, Elana E, Eric M, Tal P, Nir B (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research*, 38: W529–W533.
- Ippolito JA, Alexander RS, Christianson DW (1990). Hydrogen bond stereochemistry in protein structure and function, *J. Mol. Biol.*, 215(3):457-71.
- Jae-Min S (2005). PDB-Ligand: a ligand database based on PDB for the automated and customized classification of ligand-binding structures. *Nucleic Acids Research*, 33:238-241.
- Janin J, Miller S, Chothia C (1988). Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.*, 204(1):155-64.
- Jianghong A, Takao N, Yasushi K, Akinori S (1998). 3DinSight: an integrated relational database and search tool for the structure, function and properties of biomolecules, *Bioinformatics*, 14(2):188-195.
- Kasinos N, Lilley GA, Subbarao N, Haneef I (1992). A robust and efficient automated docking algorithm for molecular recognition. *Protein Eng.*, 5(1):69-75.
- Kitajima K, Shandar A, Samuel S, Hideo K, Shinji S, Jianghong A, Akinori S (2002). Development of a Protein-Ligand Interaction Database, ProLINT and Its Application to QSAR Analysis. *Genome Informatics*, 13:498-499.
- Loren LL, Dwyer MA, Smith JJ, Hellinga HW (2003). Computational design of receptor and sensor proteins with novel functions. *Nature*, 423:185-190.
- McDonald IK, Thornton JM (1994). Satisfying Hydrogen Bonding Potential in Proteins, *J. Mol. Biol.*, 238:777-793.
- Morris AS, Thanki N, Goodfellow JM (1992). Hydration of amino acid side chains: dependence on secondary structure. *Protein Eng.*, 5(8):717-28.
- Murzin AG, Brenner SE, Hubbard T, Chothia C (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, 247(4):536-540.
- Nicola DG, Richard MJ (2006). SitesBase: a database for structure-based protein–ligand binding site comparisons. *Nucleic Acids Research*, 34:231-234.
- Reddy AS, Amarnath HS, Bapi RS, Sastry GM, Sastry GN (2008). Protein ligand interaction database (PLID). *Comput. Biol. Chem.*, 32(5):387-90.
- Serrano L, Day AG, Fersht AR (1993). Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *J. Mol. Biol.*, 233(2):305-12.
- Smellie AS, Crippen GM, Richards WG (1991). Fast drug-receptor mapping by site-directed distances: a novel method of predicting new pharmacological leads. *J. Chem. Inf. Comput. Sci.*, 31(3):386-92.
- Smith-Gill SJ, Wilson AC, Potter M, Prager EM, Feldmann RJ, Mainhart CR (1982). Mapping the antigenic epitope for a monoclonal antibody against lysozyme. *J. Immunol.*, 128(1):314-22.
- Stuart AC, Ilyin VA, Sali A (2002). LigBase: a database of families of aligned ligand binding sites in known protein sequences and structures. *Bioinformatics*, 18(1):200-201.
- Subbarao N, Haneef I (1991). Defining topological equivalences in macromolecules. *Protein Eng.*, 4(8):877-84.
- Wallace AC, Laskowski RA, Thornton JM (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.*, 8(2): 127-134.
- Wilfred WL, Boojala VBR, John GT, Ilya NS, Philip EB (2002). CKAAPsDB: a Conserved Key Amino Acid Positions DataBase. *Nucleic Acids Res.*, 30(1):409-411.
- Zhi-Jie L, Yuh-Ju S, John R, Yong-Je C, Chwan-Deng H, Wen-Rui C, Ingrid K, John P, Ronald L, John H, Bi-Cheng W (1997). The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold, *Nat. Struct. Biol.*, 4:317-326.