PROTEINS: Structure, Function, and Genetics 52:400–411 (2003)

A Consensus-Binding Structure for Adenine at the Atomic Level Permits Searching for the Ligand Site in a Wide Spectrum of Adenine-Containing Complexes

Yosef Y. Kuttner, Vladimir Sobolev,* Alexander Raskind, and Marvin Edelman
Department of Plant Sciences, Weizmann Institute of Science, Rehovot, Israel

ABSTRACT **Attempts to derive structural features of ligand-binding sites have traditionally involved seeking commonalities at the residue level. Recently, structural studies have turned to atomic interactions of small molecular fragments to extract common binding-site properties. Here, we explore the use of larger ligand elements to derive a consensus binding structure for the ligand as a whole. We superimposed multiple molecular structures from a nonredundant set of adenosine-5′-triphosphate (ATP) protein complexes, using the adenine moiety as template. Clustered binding-site atoms of compatible atomic classes forming attractive contacts with the adenine probe were extracted. A set of atomic clusters characterizing the adenine binding pocket was then derived. Among the clusters are three vertices representing the interactions of adenine atom N6 with its protein-binding niche. These vertices, together with atom C6 of the purine ring system, complete the set of four vertices for the pyramid-like structure of the N6 anchor atom. Also, the sequence relationship for the adenine-binding loop interacting with the C2–N6 end of the conjugated ring system is expanded to include a third hydrophilic cluster interacting with atom N1. A search procedure involving interatomic distances between cluster centers was formulated and applied to seek putative binding sites in test cases. The results show that a consensus network of clusters, based on an adenine probe and an ATP-complexed training set of proteins, is sufficient to recognize the experimental cavity for adenine in a wide spectrum of ligand–protein complexes. Proteins 2003;52:400–411.**

Key words: ATP; ligand binding; atom–atom contacts; ligand design; structural similarity

INTRODUCTION

The search for conserved aspects of protein structure has often generated fundamental insights in biochemistry. Many attempts to derive structural features of binding sites were performed on proteins complexing adenosine-5′-triphosphate (ATP), an important and frequently appearing ligand in the structural Protein Data Bank (PDB).[1,2] Various sequence motifs have been proposed for the ATP-binding proteins: The A motif,[3] the P- loop,[4] the GXGXXG motif in the protein kinase family[5]; and the HSP70 protein family signature.[6] Comparative structural studies on mononucleotide-binding proteins have also been carried out. These identified various sequence motifs[7] and a variety of protein folds.[8] However, the information was not sufficient to define the ligand-binding schemes. Therefore, structural studies were initiated by several groups to extract common binding-site properties. Moodie et al.[9] described the recognition of adenylate by proteins in terms of a fuzzy recognition template, with hydrophobic residues forming a sandwich-like structure as its basis. They concluded that although certain properties of the protein–ligand interface are common for all complexes, there is no specific recognition motif in terms of particular residue–ligand interactions. To find commonalities at the residue level in hydrophobic regions of adenine-binding sites, it is necessary to consider families of proteins separately.[10]

Structural analysis of atomic interactions between the nucleotide base of ATP or guanosine-5′-triphosphate and proteins revealed the structural variety existing for purine base recognition[11] and showed that proteins having totally different folds may adopt similar recognition schemes.[12] A method was developed to search for similar local protein structures at ligand-binding sites with the use of pairwise superimposition of proteins.[13] Analyzing hydrogen-bond interactions, Denessiouk and Johnson[14] showed that 12 different fold types share a specific recognition pattern for the adenine moiety, and 8 of these have a common structural framework at the residue level for recognition of the adenosine monophosphate molecule. A computerized listing of crystallographic and theoretical data on nonbonded intermolecular interactions was developed,[15] and the generated library was used to identify interaction sites. This was achieved by superimposition of molecular fragments from several files, with placement of all contacting groups in the same system of coordinates. Cataloging the contacting groups provides a valuable resource that might be applicable for deriving a consensus structure.

Grant sponsor: V.S. received partial support from the Israel Ministry of Absorption, the Center for Absorption of Scientists.

*Correspondence to: Vladimir Sobolev, Department of Plant Sciences, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: vladimir.sobolev@weizmann.ac.il

Received 22 July 2002; Accepted 10 January 2003

Recently, with the use of multiple-ligand superimposition, the three-dimensional (3D)-distribution of binding pocket atoms around the adenine ring was illustrated, and clusters of hydrophilic and hydrophobic protein atoms were recognized in both nonredundant[16] and redundant[17] data sets. Using molecular fragment superimposition,[18,19] Rantanen et al.[20] went on to derive clusters of contacting atoms for fragments consisting of three to four atoms and produced a protein–atom/ligand-fragment interaction library. Comparable libraries have also been generated[5,21,22] from analysis of the Cambridge Structural Database. These libraries have been used to predict interactions within a binding site and to identify the ligand atom type that most probably fits at different locations within the binding site.[23]

We seek specific modes of recognition at the atomic level for larger, self-contained ligand moieties to derive a consensus-binding structure for the whole ligand. The relative positions of certain protein atoms in contact with the probe element and common to a nonredundant group of structures are calculated, and interatomic distances between key protein atoms are then applied to search for putative binding sites. There are several procedures to find cavities within protein structures (see Chakravarty et al.[24] and references therein). However, approaches to search for specific binding sites have, so far, been restricted to finding metal binding sites,[25–27] determining enzyme active sites at the residue level,[28] and predicting surface patches responsible for carbohydrate binding.[29] In this study, we employ the adenine moiety as a probe molecule and LPC software[30] as a tool for deriving the ligand–protein interface. A search procedure is formulated that successfully locates the receptor cavity for adenine-containing ligands in a wide spectrum of proteins.

MATERIALS AND METHODS

LPC Software

LPC software definitions are used throughout this article to describe ligand–protein interfaces.[30] Data provided by this tool are based on analyses of interatomic contacts.[31] The software lists the residues in contact with the ligand, thus delineating the binding site. It also defines the binding-site atoms in direct contact with specific ligand atoms and their atomic hydrophobic–hydrophilic properties. Eight classes of atoms, defined by atom type and bonding arrangement, were used in this study.[32]

Defining the Data Sets

Training set of ATP-containing entries

A list of proteins complexed with ATP and resolved to 2.2 Å or better was prepared (PDB[1,2] release, October 1999). ClustalW software[33] was used to estimate the sequence similarity among proteins in the list. If a pair of proteins had a score >30, one protein was omitted. The ligand–solvent accessibility ratio of complexed versus uncomplexed forms was determined for all structures left in the list. Entry 1b0u proved to be anomalous, with the solvent accessibility differing by more than 14 standard deviations from the average of the other entries (the adenine ring of entry 1b0u is surface located and has stacking interaction with one aromatic residue[34]); therefore, this entry was excluded. The 14 final entries in the training set are listed in Table I.

TABLE I. Training Set of Protein Complexes Containing ATP

PDB entry	Protein	Resolution (Å)
1a49	Pyruvate kinase	2.10
1a6o	Protein kinase ck2 alpha-subunit	2.10
1a82	Dethiobiotin synthetase	1.80
1atp	C-Amp-dependent protein kinase	2.20
1ayl	Phosphoenolpyruvate carboxykinase	1.80
1b8a	Aspartyl-trna synthetase	1.90
1csn	Casein kinase-1	2.00
1hck	Cyclin-dependent kinase 2	1.90
1kay	70 kd Heat shock cognate protein	1.70
1mjh	Universal stress protein family	1.70
1nsf	N-Ethylmaleimide sensitive factor	1.90
1phk	Phosphorylase kinase	2.20
1qmm	Phosphatidylinositol 3-kinase	2.20
1yag	Gelsolin	1.90

TABLE II. Test Set of Protein Complexes Containing ATP

PDB entry	Protein	Resolution (Å)
1e2q	Thymidylate kinase	1.70
1f9a	Nmn adenylyltransferase	2.00
1fmw	Myosin motor domain	2.15
1g5t	Corrinoid adenosyltransferase	1.80
1gn8	Phosphopantetheine adenylyltransferase	1.83
1hp1	5′-Nucleotidase	1.70
1j7k	DNA helicase RUVB	1.80
1jjv	Dephosphocoenzyme-A kinase	2.00
1kp2	Argininosuccinate synthetase	2.00

Test set of ATP-containing entries

All recent protein structures from 2000 and 2001 appearing in the PDB (release, June 2002) complexed with ATP and resolved to 2.15 Å or better were taken. When two entries for the same protein were encountered, the one with the highest resolution was taken (or the first released, when resolutions were equal). Entries with an additional heterogroup contacting adenine (1ee1, 1dv3, 1esv), or with adenine contacting two separate chains (1f2u), were discarded. The 9 entries of this test set are listed in Table II.

Test set of adenine-containing ligands other than ATP

An independent data set resolved to 2.0 Å or better was generated from PDB entries having an adenine-containing ligand other than ATP. In order to identify a maximal number of appropriate entries, an approach based on biological function was chosen. The PROSITE database (http://www.expasy.ch/prosite/) that relates primary structure of proteins to their function was analyzed, and

TABLE III. Test Set of Proteins Complexed with Adenine-Containing Ligands Other Than ATP

PDB Entry	Protein	Ligand	Resolution (Å)
1ads	Aldose reductase	NAD phosphate	1.65
1b4v	Cholesterol oxidase	Flavin-adenine dinucleotide	1.50
1bx4	Adenosine kinase	Adenosine	1.50
1byq	Heat shock protein 90	Adenosine-5′-diphosphate	1.50
1csc	Citrate synthase	Carboxymethyl coenzyme A	1.70
1kpf	Protein kinase C inhibitor	Adenosine monophosphate	1.50
1mmg	Myosin	Phosphothiophosphorate-adenylate	1.90
1nhk	Nucleoside diphos. kinase	Cyclic adenosine monophosphate	1.90
1zin	Adenylate kinase	Bis(adenosine)-5′-pentaphosphate	1.60
2src	Tyrosine-protein kinase	5′-Adenyly-imido-triphosphate	1.50
9ldt	Neuraminidase	Nicotinamide adenine dinucleotide	2.00

families of proteins known to interact with nucleotides were selected. The corresponding PDB entries were extracted from PROSITE records. Some additional structures were identified by scanning PDB with PROSITE consensus patterns (http://www.npsa-pbil.ibcp.fr/cgi-bin/pattern_pattinprot.pl). A small, web-driven database (http://sgedg.weizmann.ac.il/lpraskind/atpbinfam.html) was then created and used to compose an independent data set of entries. The highest resolution structure for each ligand was selected. When two entries for the same protein were encountered, the one with the highest resolution was taken (or the first released, when resolutions were equal). The final "non-ATP" test set of 11 PDB entries is listed in Table III.

Superimposition of Binding Pockets and Clustering of Atoms

Multiple structural alignment of binding sites and clustering of atoms in 3D space was carried out in the following steps:

1. For each entry in the input list of PDB entries, protein atoms in contact with the ligand were found using LPC software,[30] and their coordinates assigned to a file.
2. Molecular structures from these files were superimposed using the adenine moiety as template. The adenine rings from each structure were placed in a common coordinate system, such that C1 atoms were at the origin, N2 atoms at the X ordinate, and C7 atoms on the XY plane.
3. A list of all the protein atoms in contact with the ligand was created, and for each atom in this list, the number of atomic neighbors within a designated threshold distance, and contacting the same ligand atom, was recorded.
4. Protein atoms were then sorted by decreasing number of neighbors.
5. The atom with the most neighbors, and all its neighbors within a given threshold distance, were deleted from the table and outputted to a file containing the coordinates of atoms for a given cluster.
6. Step 5 was repeated to obtain additional clusters, until no further atoms remained in the table.

Finally, in each cluster, only atoms forming attractive contact with the ligand were accepted, and only one atom (if any), that closest to the cluster center, was kept for each PDB entry. The clustering procedure defines a "cluster type" based on the identity of the most frequent, non-neutral atom class contributing to the cluster. Thus, a cluster is composed of accepted atoms from one or more classes,[32] where all atoms within one cluster form attractive contacts with the ligand. After preliminary trials, we settled on a maximum cluster radius of 1.5 Å, and accepted clusters as representative if they consisted of 50% or more of the PDB entries under analysis.

RESULTS AND DISCUSSION

Defining a Cluster

Using procedures described in the Materials and Methods section, 12 clusters for the adenine moiety were extracted from our nonredundant data set of 14 ATP-containing files. These clusters and the atom types within each cluster are listed in Table IV. Note that we accepted donor atoms (class III) from PDB entries 1mjh and 1ayl as bona fide members of cluster 1 (cluster type, Hydrophobic). This is because the aromatic rings of adenine can act as hydrogen-bond acceptors[35–39]; therefore, in cluster 1, both hydrophobic and donor atoms can form attractive contacts with the conjugated system of the adenine rings, and both classes may participate. Similarly, we accepted hydrophilic atoms (classes I and III) in Hydrophobic clusters 3 and 4, and an aromatic atom (class V) in Hydrogen Bond Acceptor cluster 10 (Table IV). However, backbone oxygen (a hydrogen-bond acceptor) of Asp176 from entry 1a82, although within the radial range of cluster 2, would form a repulsive contact with the electron cloud of the aromatic rings of adenine. Thus, this atom was excluded and an empty cell is found for cluster 2 at the 1a82 slot in Table IV. For the same reason, we excluded the oxygen atom of Val177, entry 1a82 from cluster 7. Furthermore, atom CB of Val882A in entry 1qmm was excluded from cluster 8 because of repulsive contact with atom N1 of adenine, whereas atom CB of Phe113 in entry 1a6o was excluded from cluster 10 because of repulsive contact with N6. Thus, the sets of potentially compatible atom classes are: I, III,

TABLE IV. Cluster Properties

PDB entry	Cluster number											
	1	2	3	4	5	6	7	8	9	10	11	12
	Atom class[a]											
1phk	IV	IV	IV	—	IV	IV	IV	III	IV	II	IV	II
1hck	IV	IV	IV	—	V	IV	IV	III	IV	II	IV	II
1atp	—	IV	IV	IV	V	IV	IV	III	IV	II	—	II
1csn	IV	IV	—	IV	IV	IV	IV	III	—	II	IV	—[b]
1mjh	III	IV	IV	I	—	IV	IV	III	—	II	—	II
1a82	IV	—[b]	IV	IV	V	—	—[b]	III	IV	II	IV	II
1a6o	IV	V	IV	III	IV	—	IV	—	IV	—[b]	—	VI
1ayl	III	VI	—	—	IV	—	—	—	IV	I	IV	II
1a49	VI	VI	IV	—	—	—	IV	—	VI	V	IV	—
1kay	—	IV	IV	VI	—	VI	—	I	—	—	—	—
1qmm	IV	—	—	IV	—	IV	—	—[b]	V	—	IV	—
1b8a	VI	—	III	—	—	IV	IV	III	—	—	—	—[b]
1yag	—	IV	IV	I	VI	VI	—	—	—	—	—	—
1nsf	IV	IV	—	IV	IV	—	—	—	—	—	—	—

A filled cell indicates that the entry contributes a member to the cluster, an empty cell (—) indicates that it does not. Shading indicates cluster type, which is assigned based on the most frequent class in the cluster: white, Hydrophobic (IV); light gray, Hydrogen Bond Donor (III); dark gray, Hydrogen Bond Acceptor (II).

[a]Atom classes[32]: *I, Hydrophilic:* N and O that can donate and accept hydrogen bonds; *II, Acceptor:* N or O that can only accept a hydrogen bond; *III, Donor:* N that can only donate a hydrogen bond; *IV, Hydrophobic:* CI, Br, I, and all C atoms that are not in aromatic rings and do not have a covalent bond to an N or O atom; *V, Aromatic:* C in aromatic rings irrespective of any other bonds formed by the atom; *VI, Neutral:* S, F, P, and metal atoms in all cases, and nonaromatic C atoms that have a covalent bond to at least one atom from class I, or two or more atoms from class II or III; *VII, Neutral donor:* Nonaromatic C atoms that have a covalent bond with only one atom of class III; *VIII, Neutral acceptor:* Nonaromatic C atoms that have a covalent bond with only one atom of class II.

[b]Although an atom is within the cluster radius, it is not accepted as a cluster member due to unfavorable interaction (see text).

IV, V, VI, and VII for Hydrophobic clusters; I and III for Hydrogen Bond Donor clusters; and I, II, V, VI, and VIII for Hydrogen Bond Acceptor clusters. However, for analysis of the clusters in Table IV, only the compatible atom classes actually found were included (for cluster 1, only classes III, IV, and VI were accepted, etc.).

There is a total of 352 nonhydrogen atoms in contact with adenine within our data set. Of these, 138 are in the defined cluster space, and 105 were accepted in the 12 clusters shown in Table IV. The 33 that were not included represent either repulsive contacts or multiple atoms from the same entry. The average density in the cluster space is 2.4 atoms per Å^3 (range 1.4–4.5 atoms per Å^3) and 0.7 atoms per Å^3 in the nonclustered space. The total average density for the 352 contacting atoms is approximately one atom per Å^3. Our clustering procedure captured 40% of the atoms in contact with the adenine moiety. Concerning the remaining 60%, the exclusion of a portion can undoubtedly be ascribed to imperfect definition of clusters, incompleteness of the cluster set, insufficient resolution of crystal structures, or the natural diversity of the adenine-binding pocket in our wide-spectrum training set (Table I). However, a significant portion of the nonclustered, contacting atoms is likely to have an intrinsically distinct role—that of providing a matrix for the clusters. We speculate that whereas clustered atoms impart specificity, nonclustered ones impart stability: in this case, an energetically not-insignificant complementary shell to the planar, conjugated ring system of adenine.

Cluster Description

The 3D location of the clusters around the adenine ring is depicted in a perspective projection from two vantage points in Figure 1. In a frontal view, perpendicular to the plane of the rings (Fig. 1, upper panel), all clusters appear well delineated except clusters 1 and 3, whose closest member atoms are 0.7 Å apart. There are eight Hydrophobic clusters (clusters 1–7 and 11) distributed above and beneath the plane of the conjugated ring system, and four clusters, two (clusters 8 and 10) strongly and two (clusters 9 and 12) weakly hydrogen-bonded to the distal end of the adenine moiety (at atoms N6, N1, and C2). In Figure 1, lower panel, the plane of the ring system is rotated 90°, such that the four clusters (clusters 1, 3, 4, and 11) beneath the plane in Figure 1, upper panel, are now more spread out.

Multiple contacts

Table V shows the dominant adenine atom(s) contacted by each cluster, indicating the general position of a cluster relative to the adenine moiety. The full set of such contacts can be found at http://sgedg.weizmann.ac.il/recognition/adenine. The hydrophobic clusters have a total of 82 nonhydrogen protein atoms in contact with adenine: 90% are derived from amino acid sidechains, and more than 90% of these are carbon atoms. Conversely, a preponderance (19 of 23) of hydrophilic cluster atoms are from the protein backbone, in agreement with the description by Denessiouk et al.[17] Thus, although the adenine pocket is basically hydrophobic, hydrophilic interactions can be

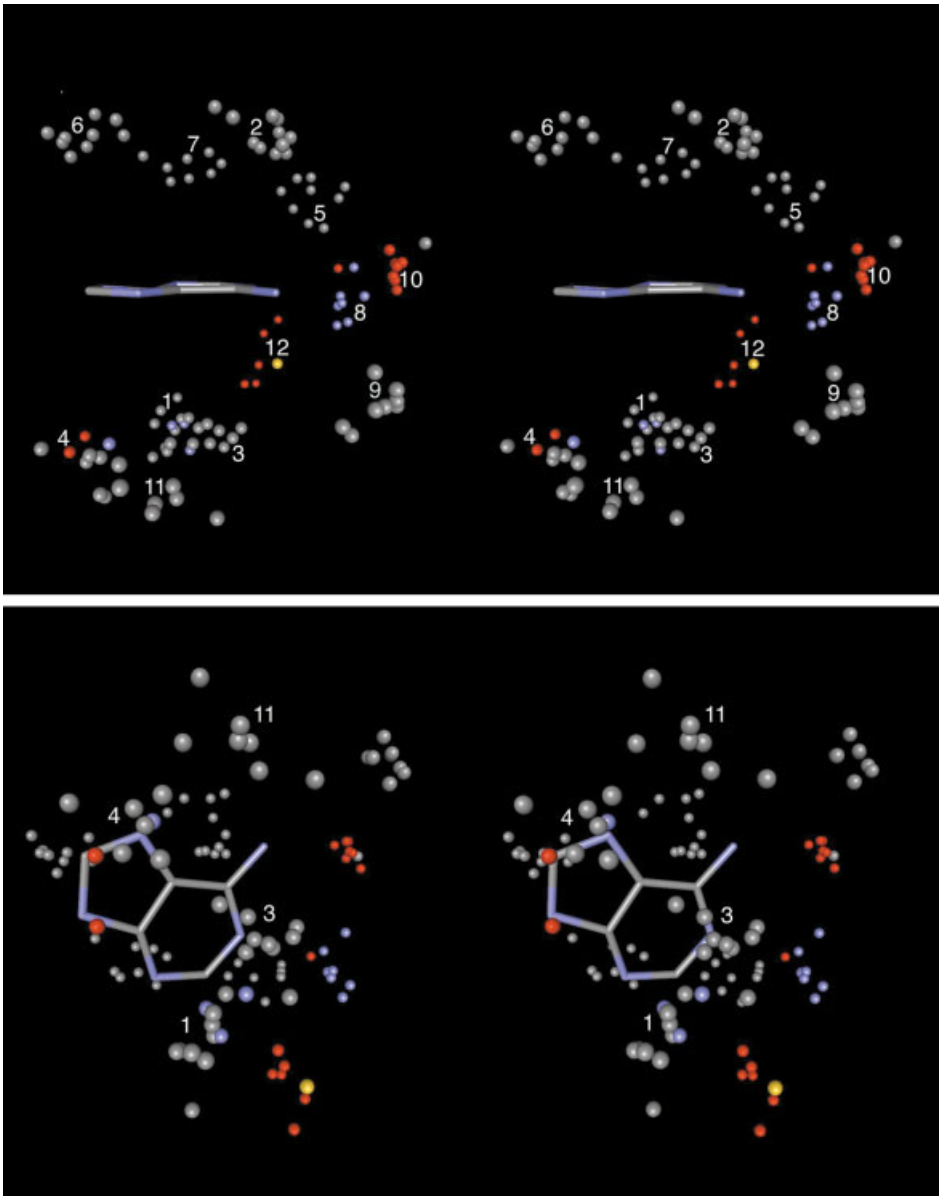


Figure 1.

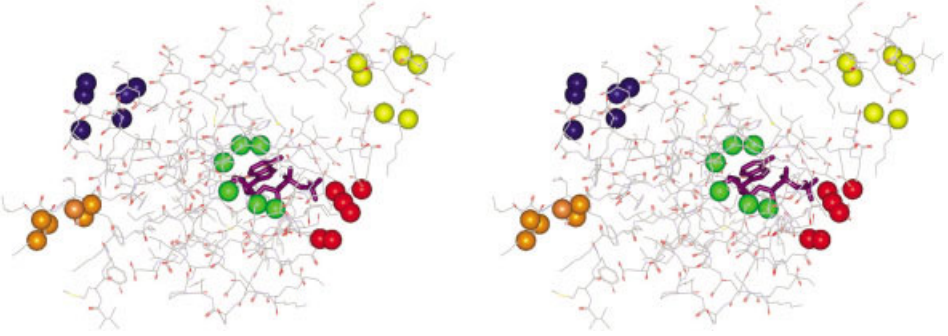
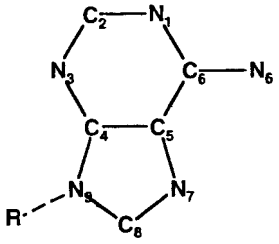


Figure 2.

TABLE V. Atoms of Adenine in Contact with Clustered Atoms of Proteins[a]

Adenine		Cluster No.											
Atomic designation	Atom	1	2	3	4	5	6	7	8	9	10	11	12
	N1		+			+		+	+				
	C2	+				+		+					
	N3	+						+					
	C4	+					+						
	C5			+			+						
	C6		+	+									
	N6		+	+						+	+		
	N7				+		+					+	
	C8				+		+						
	N9						+						
No. of contacts[b]		3.3	3.3	2.9	1.6	1.3	3.8	2.9	1.3	1.0	1.0	1.1	1.3
Z_c(Å)[c]		−3.8	+3.5	−3.7	−3.4	+2.4	+3.8	+3.5	−0.6	−2.2	+0.3	−3.6	−2.5

[a]Adenine atoms marked as (+) collectively contact >70% of the protein atoms in that cluster.
[b]Average number of contacts of a cluster atom with atoms of the adenine moiety.
[c]Distance of the cluster center from the plane of the adenine rings. Minus (−) or plus (+) indicates the position of the cluster relative to the conjugated ring system, as pictured in Figure 1, upper panel.

formed with polar backbone atoms. Hence, atomic (rather than residue) recognition imparts specificity to adenine binding. Hydrophobic cluster atoms often form multiple (up to six) atomic contacts when positioned above or beneath the plane of the adenine ring system, whereas atoms from all four hydrogen-bonded clusters form basically a single contact with adenine. The average number of ring contacts per atom per cluster is given at the bottom of Table V.

Positioning

Seven hydrophobic clusters (clusters 1–4, 6, 7, and 11) are located 3.4–3.8 Å from the plane of the adenine rings (Z_c, Table V). These values are close to the sum of van der Waals radii for C—C and C—N interactions, indicating that these clusters are situated almost directly above and beneath the rings (cf. Fig. 1, upper panel). On the other hand, hydrophobic cluster 5, which interacts with atoms C2 and N1, is located 2.4 Å above the ring plane, indicating an oblique positioning. Interestingly, cluster 12 geometrically counterbalances cluster 5 (cf. Fig. 1, upper panel), being is positioned 2.2 Å beneath the ring plane. Cluster 12 is composed mainly of oxygen atoms in contact with adenine atom C2. Such CH…O contact is generally accepted as a weak hydrogen bond.[10,40–44]

Hydrophobic cluster 9 is also obliquely positioned (distance, −2.2 Å) compared to the ring plane. The atoms in this cluster interact solely with hydrophilic atom N6 of adenine (Table V). At first glance, this would appear to indicate a repulsive contact. However, Luisi et al.[45] showed that the amino group of adenine has the highest propensity among nucleobases to accept a hydrogen bond from a CH group, and they consider this contact to be attractive. The positions of clusters 9 and 10 relative to the C6—N6 covalent bond of adenine show characteristic pyramidalization of the amino group, supporting the assumption that the N6 atom of the adenine moiety contacting cluster 9 is indeed acting as a hydrogen-bond acceptor. Further backing can be gleaned from a recent report[10] of a second, strong hydrogen-bond location interacting with adenine atom N6 in addition to cluster 10. Essentially, all four vertices of a pyramid-like structure representing the interactions of atom N6 with the adenine ring system and with its protein binding niche can now be described (namely, atom N6 at the center, and atom C6, cluster 9, cluster 10, and the strong hydrogen-bond location of Cappello et al.[10] at the vertices). Thus, N6 appears to be an anchor atom, interacting with its proteinaceous surrounding by up to three hydrogen bonds: two strong and one weak.

The atomic positions forming hydrogen bonds with N6 and N1 of adenine have been well described.[9–11,14,20] Clusters 10 and 8 in our study fit into those positions, forming strong hydrogen bonds with atoms N6 and N1 of adenine, respectively. These two clusters are situated almost on the plane of the ring (Fig. 1, upper panel).

Fig. 1. Positions of clusters around the adenine moiety. The adenine moieties of the 14 structures in the ATP-containing training set were superimposed and clusters of atoms extracted as described in Materials and Methods section. Cluster numbers follow those in Table IV. Carbon atoms are colored gray; oxygen atoms, red; nitrogen atoms, blue; and sulphur atoms, yellow. The pictures were prepared with WebLab Viewer software (Molecular Simulations, Inc.), using a perspective projection view. **Upper panel**, frontal view perpendicular to the plane of the adenine rings. **Lower panel**: plane of ring system rotated 90° such that clusters 1, 3, 4, and 11 are now extended.

Fig. 2. Positions of the five putative binding sites of entry 1mjh. Training set entry 1mjh is predicted to have five putative binding sites (cf. Table VI). Four of the sites are located at the convex surface, whereas the remaining one (colored green) is immersed in the protein core. This site is the correct prediction. The ATP ligand at the experimental position is shown in purple. Analysis was performed by extracting the coordinates for the putative sites from the PDB file and highlighting their positions in WebLab Viewer (MSI).

Kobayashi and Go[12] described a structurally conserved segment of four residues, two of whose backbone atoms (from the second and forth residue) form hydrogen bonds with the adenine moiety of ATP. In a more extensive study,[14] this was corroborated and, additionally, it was found that in some proteins, a single residue forms two hydrogen bonds with N6 and N1 of the adenine moiety. Our results now refine these adenine-binding motifs to include all three hydrophilic clusters interacting with the N6, N1, and C2 atoms of the ligand. We find that clusters 10, 8, and 12, respectively, have a tendency to be populated by residues whose sequence along the polypeptide chain is X, X+2, X+2 (e.g., in protein 1hck, the cluster atoms come from residues 81, 83, and 83, respectively). In these cases, backbone oxygen from the first residue (X), and backbone nitrogen from the second (X+2), hydrogen-bond with adenine atoms N6 and N1, respectively, whereas the backbone oxygen from the second residue (X+2) hydrogen-bonds with atom C2. In our training set of ATP-containing files (Table I), this pattern is discerned by LPC analysis in entries 1hck, 1atp, 1phk, 1a82, 1csn, and 1qmm. Two additional entries in the ATP-containing training set (1mjh and 1b8a) follow a related adenine-binding motif pattern of X, X, X−2 for the respective amino acid residues of the three aforementioned clusters. This latter pattern, however, is more characteristic of flavin-adenine dinucleotide complexed with protein (13 of 21 nonredundant files). The data set and results for these complexes can be found at http://sgedg.weizmann.ac.il/recognition/adenine.

Searching for Adenine-Binding Sites

Search procedure

A search procedure was formulated to test the degree of commonality in atomic distribution within adenine-binding sites from different proteins. Atoms forming a binding site will, to a first approximation, be surface-exposed in an apo-protein. Therefore, the search procedure initially finds all solvent-accessible atoms in the protein structure, using a probe sphere of van der Waals radius 1.4 Å. Further relevant parameters are intercluster distances (i.e., the maximal and minimal distances, R_{ab}^{min} and R_{ab}^{max}, between atoms belonging to different clusters a and b) and the minimal number (n) of atoms in a group. Briefly, the search procedure seeks a group of n atoms, belonging to N clusters with specified intercluster distances, and with all atoms coming from compatible classes. The search procedure is computationally similar to pairwise structural alignment of the derived, consensus binding site and the crystallized protein. Although pairwise structural alignment is a complex procedure,[46] when the consensus binding site consists of a small number of points (say, 10–20), exact solution(s) can be obtained in reasonable computer time.

Intercluster distances (R_{ab}^{min} and R_{ab}^{max}) were sought either with or without considering from which PDB entries which the atoms derive. Preliminary experimentation showed that a smaller number of false positives are encountered with the former, and this method was adopted. Distances between atoms belonging to clusters a and b are calculated, with both atoms coming from the same PDB entry. Then, the corresponding maximum and minimum values between cluster atoms are chosen. Next, geometric restrictions are set: If the search procedure finds two atoms, one putatively belonging to a cluster above the plane of the adenine rings and the other below, there must be sufficient cavity space at that location in the protein structure for the adenine rings to fit between the contacting atoms. To accomplish this, we calculated the distance between the geometric center of the two cluster atoms and the nearest protein atom for every PDB structure and every pair of clusters in Table IV. For every pair of clusters, "minimal distance" was assigned as a parameter. During the search procedure, two atoms in a protein are considered as belonging to these clusters only if the distance from their geometric center and nearest protein atom is larger than this "minimal distance."

Searching within the training set

The training set consisted of 14 nonredundant, high-resolution entries of ATP-complexed proteins (Table I). We found that by considering all 12 clusters, several thousand groups of 4 or 5 atoms score positive in most entries. On the other hand, a group of 7 atoms never scored positive for all the entries of the data set. We therefore settled on 6 atoms for search purposes, one from each of 6 of the 12 clusters. The first acceptable group of 6 atoms is taken as an initial binding site. Any following group either forms a new putative site or refines an existing one. A group will refine an existing site if the distance between its geometric center and that of the site is less than 10 Å. The center of a site is dynamically defined as the geometric center of all the groups forming it. The 10 Å threshold was chosen for two reasons. First, the shortest distance between geometric centers of coplanar cavities is the sum of the van der Waals radii from the center of the first cavity to the nearest layer of atoms forming the cavity wall (~3.5 Å), the thickness of two monolayer walls (~3.5 Å), and the distance from the wall of the second cavity to its geometric center (~3.5 Å). Second, when the distance between the geometric centers of two groups of 6 atoms was less than 10 Å, we found that the sets of residues for these groups usually overlap by at least one residue. Any acceptable group of 6 atoms defines a "hit". A correct site has at least one hit containing at least one atom in contact with adenine in the experimental structure. All other sites are defined as false positive.

The search results for the training set are summarized in Table VI. The number of "Cluster atoms" present in each PDB entry is shown. For some entries, 11 of 12 atoms are present, resulting in a large number of "Expected hits" at the correct site. For others, the number of atoms present is only 4 or 5, resulting in no expected hits at all. To overcome the latter difficulty, we searched for additional, potentially relevant atoms not included by the clustering procedure. A significant amount of protein atoms in the cluster space are located more than 4 Å from their closest ligand atom. Protein atoms such as these can be screened by more closely positioned ones that, although not in the

TABLE VI. Search Results for Training Set of ATP-Complexed Proteins

PDB entry	Cluster atoms[a]	Expected hits[b]	Total hits[c]	Best-hit distance[d] (Å)	Putative sites: Total no.[e]	Narrowing of false positives by: Chain size[f]	Visual[g]	Most hits[h]	Hits + visual[i]
1phk	11	462	534	0.1	9	4	1	0	0
1hck	11	462	518	0.1	12	9	5	0	0
1atp	10	210	471	0.1	11	9	2	0	0
1csn	9	84	230	0.2	8	4	0	0	0
1mjh	9	84	179	0.2	5	4	0	0	0
1a82	9	84	143	0.1	5	4	0	0	0
1a6o	8	28	103	0.2	12	7	0	0	0
1ayl	7	7	92	0.5	18	14	5	0	0
1a49	7	7	51	0.4	13	9	5	0	0
1kay	5	0	71	1.0	12	8	5	0	0
1qmm	5	0	239	8.7[j]	25	22	6	0	0
1b8a	5	0	60	2.4	17	12	2	7–12[k]	1
1yag	5	0	51	2.4	16	9	1	1	0
1nsf	4	0	56	1.2	10	9	2	0	0

[a]Number of clustered atoms at the authentic site (derived from Table IV).
[b]Number of hits (K) theoretically expected for 6 atoms contacting adenine at the correct site given the number of cluster atoms (k) in the specified PDB entry: $K = k!/(6!\,(k-6)!)$.
[c]Total number of hits found for all putative sites.
[d]Distance between the center of the experimental binding site and the center of atoms forming the closest hit.
[e]Total number of putative sites found.
[f]Number of false positives remaining following a chain-length restriction of at least 15 residues between distal amino acids of a putative site.
[g]Number of false positives remaining following visual elimination of convex surface sites using WebLab Viewer (MSI).
[h]Number of false positives remaining after applying "highest frequency of hits at a binding site" as a limiting parameter.
[i]Number of false positives remaining after applying "Visual" and "Most hits" techniques in combination.
[j]In spite of the large distance, three atoms of the best hit are in contact with adenine.
[k]Identical frequency of hits was found for multiple putative sites.

cluster, are in contact with the ligand target atom (cf. Fig. 2 in Sobolev and Edelman[31]). In such instances, the distant atom does not appear in the clustering data of Table IV (because the clustering procedure requires contact with the ligand atom and this, by definition, is provided by the closer atom), but is recognized by the search procedure (because cluster space and atom type are the parameters scored in the search). Such additional atoms allowed the search for groups of 6 atoms to proceed in entries with less than 6 clusters.

The total number of hits ("Total hits" column) exceeded the expected hits for each entry in the training set (Table VI). This is readily explained, because only sets of atoms in which all 6 members are in contact with adenine are represented in "Expected hits." On the other hand, in "Total hits," additional sets populate false-positive cavities, or even the correct cavity, but have less than 6 atoms in contact with adenine. Critically, the data in Table VI show that the correct experimental cavity for adenine was found in all 14 cases. This is indicated by the small distance between the center of the experimental site for each entry and the closest hit from the search results ("Best-hit distance" column). The only seeming exception is entry 1qmm. However, in spite of the relatively large distance (8.7 Å) of the best hit from the center of the experimentally correct cavity, 3 of 6 search atoms proved to be in contact with adenine.

Searching with the test sets

The search procedure was applied to a nonredundant, high-resolution test set composed of all the suitably resolved structures complexed with ATP that appeared in the PDB following completion of our training set (Table II). The results are summarized in Table VII. With the exception of one case, the experimentally correct cavity for adenine was found for all entries (cf. "Best-hit distances" column). In the exception (entry 1hp1), the adenine ring proved to be stacked between two residues of phenylalanine.[47] A review of the atom class data in Table IV reveals that an aromatic-ring carbon atom (class V) occurred in only 3 of the 9 Hydrophobic clusters of the training set (namely, clusters 2, 5, and 9). Thus, the search procedure accepted aromatic input only for those three Hydrophobic clusters. Relaxation of the procedure to include aromaticity for all Hydrophobic clusters readily revealed the experimental binding site for entry 1hp1, with a best-hit distance of 3.6 Å, whereas the total number of putative sites increased only moderately from 18 to 21 (cf. Table VII).

We note that inclusion of hydrophilic atoms in Hydrophobic clusters 1, 3, and 4, and aromatic atoms in Hydrogen Bond Accepter cluster 10 (as described for the training set; Table IV), proved to be significant for the search procedure in the ATP test set as well (Table VII). Without such inclusion, the correct site was not found in 2 of 9 ATP test structures: 1jjv and 1e2q. Indeed, Obmolova et al.[48] con-

TABLE VII. Search Results for Test Set of ATP-Complexed Proteins

PDB entry	Total hits[a]	Best-hit distance[b] (Å)	Putative sites				
				Narrowing of false positives by:			
			Total no.[c]	Chain size[d]	Visual[e]	Most hits[f]	Hits + visual[g]
1e2q	23	0.8	7	3	0	0	0
1f9a	42	2.6	5	2	1	0	0
1fmw	144	1.6	27	16	6	3	2
1g5t	7	3.3	3	1	0	0	0
1gn8	23	5.0	7	4	1	6	1
1hp1	65	10.0[h] (3.6)[i]	18 (21)[i]	15 (16)[i]	2 (2)[i]	n.f.[h] (1)[i]	n.f.[h] (0)[i]
1j7k	45	2.2	10	5	2	0–1	0
1jjv	39	1.5	9	5	2	5–6	2
1kp2	49	6.9[j]	15	8	3	4–6	0–2

[a]See Table VI[c].
[b]See Table VI[d].
[c]See Table VI[e].
[d]See Table VI[f].
[e]See Table VI[g].
[f]See Table VI[h].
[g]See Table VI[i].
[h]Correct site not found (n.f.) using the standard search procedure (no atoms in contact with adenine in best hit).
[i]Expansion of the search function to include aromaticity in all Hydrophobic clusters (see legend, Table IV) revealed the experimental binding site.
[j]In spite of the relatively large distance, two atoms of the best hit are in contact with adenine.

cluded that Arg140A in entry 1jjv does not determine base specificity but is important for formation of the ATP pocket, and the same holds for Arg143A of entry 1e2q.

The search procedure derived for the adenine moiety of ATP-binding proteins was, in addition, applied to an independent high-resolution set of PDB entries, each member of which is complexed with an adenine-containing ligand other than ATP (Table III). These search results are summarized in Table VIII. Two of the 11 test structures of this data set contain more than one unique adenine moiety complexed to proteins: in entry 1bx4, two molecules of adenosine, and in entry 1zin, a single molecule of bis(adenosine)-5′-pentaphosphate. Thus, this test set provides 13 adenine-binding sites for the search. The experimental binding site was found for 12 of the 13 adenines using the standard search procedure, and for one of the ligands in 1bx4 following inclusion of aromaticity for all Hydrophobic clusters ("Best-hit distance" column). Thus, consensus clusters and a search procedure based on an ATP-complexed training set are sufficient to recognize the experimental cavity for adenine in a wide spectrum of ligand–protein complexes containing that purine.

For both test sets, the best hit at each correct binding site was checked with LPC software[30] and found to have at least one atom in contact with adenine. Hence, the test results show that the correct experimental cavity for adenine was discovered in all cases.

Narrowing the Number of Putative Binding Sites

Whereas the correct binding cavity for adenine was indeed discovered in all cases examined, the data also show that, all in all, about 400 putative binding sites were scored. In fact, in every case, more than one putative site was counted (Tables VI–VIII, "Total No." column) and, on average, in addition to the correct one, 10 false positives were registered. For a search procedure to be effective, it ought to find the actual binding cavity with a minimal number of false-positive sites (preferably, none). Several methods were thus surveyed in an attempt to narrow the number of false positives. The results are summarized below and in Tables VI–VIII. The data sets for these tables can be found at http://sgedg.weizmann.ac.il/recognition/adenine.

First, we experimented with reducing the number of clusters in the training set. Searching with 6 of 8 rather than 6 of 12 clusters, a grouping was found (namely, clusters 1–5, 8, 11, 12) that, on average, reduced the number of false-positive sites by 59% and the total number of hits by 89%, while retaining the experimental binding site in each training set entry. However, our search with 6 of 8 (or 6 of 9) clusters failed to find the experimental binding site for several entries of the test sets, and we abandoned this approach.

Next, we sought a common chain-length minimum that satisfied all correct sites in the training and test sets. Restrictions were set for the distal amino acids of the binding domain. Inspection of the amino acid chain encompassing the correct binding site for all members of the three data sets revealed that it is at least 15 residues long. When this empiricism is introduced as a constraint (see "Chain size" column), the number of false positives is moderately reduced, on average by 22% in the ATP-containing training set (Table VI), 27% in the ATP-containing test set (Table VII), and 23% in the non-ATP adenine-containing test set (Table VIII). However, in no case was the number of false positives reduced to nil.

In a different approach, preliminary observations showed a high percentage of bogus binding sites on the convex surface of the protein. We attempted to discriminate visually between these and putative binding sites partly

TABLE VIII. Search Results for Test Set of Proteins Complexed with Adenine-Containing Ligands Other Than ATP

PDB entry	Total hits[a]	Best-hit distance[b] (Å)	Putative sites				
				Narrowing of false positives by:			
			Total no.[c]	Chain size[d]	Visual[e]	Most hits[f]	Hits + visual[g]
1ads	47	5.4[i]	8	6	0	0	0
1b4v	65	2.5	18	12	1	3	0
1bx4[h]	72	4.8	11	8	1	0–1	0–1
	72	13.0[j] (8.2[i])[k]	11 (12)[k]	8	1	n.f.[j] (10)[k]	n.f.[j] (1)[k]
1byq	40	2.7	9	7	0	1	0
1csc	68	8.1[i]	12	8	3[l]	0	0
1kpf	49	0.8	3	2	1	0	0
1mmg	103	2.7	26	22	3	2	0
1nhk	31	6.3[i]	6	2	0	0	0
1zin[h]	28	3.2	5	2	0	2	0
	28	1.3	5	2	0	0	0
2src	94	0.3	16	12	4	0	0
9ldt	61	3.0	11	8	1	3–4	1

[a]See Table VI[c].
[b]See Table VI[d].
[c]See Table VI[e].
[d]See Table VI[f].
[e]See Table VI[g].
[f]See Table VI[h].
[g]See Table VI[i].
[h]These entries have two unique adenine moieties complexed with their proteins; in 1bx4, two molecules of adenosine; in 1zin, a single molecule of bis(adenosine)-5′-pentaphosphate.
[i]In spite of the relatively large distances, at least one atom of the best hit is in contact with adenine.
[j]Correct site not found (n.f.) using the standard search procedure (no atoms in contact with adenine in best hit).
[k]Expansion of the search function to include aromaticity in all Hydrophobic clusters revealed the experimental binding site.
[l]The correct binding site was not retained in this one case.

buried in the protein core that might accommodate an adenine moiety. Figure 2 reveals the potential of this method: In the case of training set entry 1mjh, the number of false positives was reduced from four to nil. Each entry in each data set was analyzed. The procedure retained the correct binding site in all cases except one (entry 1csc, Table VIII). The atomic coordinates for all putative sites for an entry were extracted, highlighted, and overlaid on a 3D representation of the protein using WebLab Viewer (MSI). Rotation about all axes was used to discriminate convex surface sites from partly buried ones. When convex surface sites are eliminated in this fashion, the number of false positives ("Visual" column) is reduced, on average, by 79% in the ATP-containing training set (Table VI), 82% in the ATP-containing test set (Table VII), and 88% in the non-ATP adenine-containing test set (Table VIII). In over one fourth of the cases, the number of false positives was reduced to nil.

In yet another approach, we noticed that in a majority of cases, the search procedure scored a larger number of hits at the correct binding-site cavity than at any of the incorrect ones. Parenthetically, we note that if sufficient hits exist in the correct cavity, their positional density might be useful to home in on the precise location of the correct binding position. With use of the highest frequency of hits at a binding site as a limiting parameter (see "Most hits" column), the number of false positives is reduced, on average, by 92–95% in the ATP-containing training set (Table VI), 76–80% in the ATP-containing test set (Table VII), and 80–81% in the non-ATP adenine-containing test set (Table VIII). In almost two thirds of the cases, the number of false positives was reduced to nil.

Finally, by combining the "Visual" and "Most hits" approaches ("Hits + visual" column), the number of false positives is reduced, on average, by >99% in the ATP-containing training set (Table VI), 93–95% in the ATP-containing test set (Table VII), and 97–98% in the non-ATP adenine-containing test set (Table VIII). In over three fourths of the cases, the number of false positives was reduced to nil, and in no case was it greater than two, when these two approaches were used in concert.

Currently, the consensus-binding structure approach has been applied only to rigid (adenine; this work) or almost-rigid (ribose)[49] ligand elements. In principle, this approach may be applied to rigid ligands, or ligands having a few well-defined conformations, for which sufficient cases are documented in the structural databases. With regard to flexible ligands, a two-step procedure would apply: Consensus-binding sites for individual rigid elements (e.g., the adenine, ribose, and phosphate moieties of ATP) can be derived separately, after which a search procedure accounts for the relative positioning of the elements. Such a search procedure might exploit a consensus-binding site obtained by different methods.[16,50] Extension of our work to flexible ligands is currently under development.

CONCLUSIONS

Our objective in this study was to determine whether a consensus network of atomic coordinates, derived from the

structural database, could be extracted and applied toward recognizing the binding-site cavity for a given heterogroup in any of its diverse protein complexes. The results show that for a frequently appearing PDB ligand element such as adenine, this is indeed feasible. A consensus-binding structure for adenine can be scored to predict the correct binding cavity within the protein volume with acceptable fidelity (>90%). As a consequence, it is likely that if a protein is known, or suspected, to be a target for a heterogroup containing adenine, our search procedure will allow resolution of the binding site. Nevertheless, our current procedure probably would not allow us to exclude nontarget proteins with much fidelity. This remains a challenge.

ACKNOWLEDGMENTS

We thank Dr. B. McConkey, R. Najmanovich, E.Eyal, and M. Babor for helpful discussions.

REFERENCES

1. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: A computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
2. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
3. Walker J, Saraste M, Runswick M, Gay N. Distantly related sequences in a- and b- subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and common nucleotide binding fold. EMBO J 1982;1:945–951.
4. Saraste M, Sibbald PR, Wittinghofer A. The P-loop—a common motif in ATP-binding and GTP-binding proteins. Trends Biochem Sci 1990;15:430–434.
5. Hanks SK, Quinn AM, Hunter T. The protein-kinase family—conserved feartures and deduced phylogeny of the catalytic domains. Science 1988;241:42–52.
6. Bairoch A, Bucher P, Hofmann K. The PROSITE database, its status in 1997. Nucleic Acids Res 1997;25:217–221.
7. Traut TW. The functional and consensus motifs of 9 types of peptide segments that form different types of nucleotide-binding sites. Eur J Biochem 1994;222:9–19.
8. Schulz G. Binding of nucleotides by proteins. Curr Opin Struct Biol 1992;2:61–67.
9. Moodie SL, Mitchell JBO, Thornton JM. Protein recognition of adenylate: An example of a fuzzy recognition template. J Mol Biol 1996;263:486–500.
10. Cappello V, Tramontano A, Koch U. Classification of proteins based on the properties of the ligand-binding site: The case of adenine-binding proteins. Proteins 2002;47:106–115.
11. Nobeli I, Laskowski RA, Waldar WSJ, Thornton JM. On molecular discrimination between adenine and guanine by proteins. Nucleic Acids Res 2001;29:4294–4309.
12. Kobayashi N, Go N. ATP binding proteins with different folds share a common ATP-binding structural motif. Nat Struct Biol 1997a;4:6–7.
13. Kobayashi N, Go N. A method to search for similar protein local structures at ligand-binding sites and its application to adenine recognition. Eur Biophys J 1997b;26:135–144.
14. Denessiouk KA, Johnson MS. When fold is not important: A common structural framework for adenine and AMP binding in 12 unrelated protein families. Proteins 2000;38:310–326.
15. Bruno IJ, Cole JC, Lommerse JP, Rowland RS, Taylor R, Verdonk ML. IsoStar: A library of information about nonbonded interactions. J Comput Aided Mol Des 1997;11:525–537.
16. Kuttner YY, Babor M, Edelman M, Sobolev V. Structural commonality in protein binding sites for ATP. In: El-Mabrouk, N, Lengauer T, Sankoff D, editors. Currents in computational molecular biology. Montreal: CRM; 2001. p 91–92.
17. Denessiouk LA, Rantanen VV, Johnson MS. Adenine recognition: A motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins. Proteins 2001;44:282–291.
18. Klebe G. The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. J Mol Biol 1994;237:212–235.
19. Laskowski RA, Thornton JM, Humblet C, Singh J. X-site: Use of empirically derived atomic packing preferences to identify favourable interaction regions in the binding sites of proteins. J Mol Biol 1996;259:175–201.
20. Rantanen VV, Denessiouk KA, Gyllenberg M, Koski T, Johnson MS. A fragment library based on Gaussian mixtures predicting favorable molecular interactions. J Mol Biol 2001;313:197–214.
21. Verdonk ML, Cole JC, Taylor R. SuperStar: A knowledge-based approach for identifying interaction sites in proteins. J Mol Biol 1999;289:1093–1108.
22. Verdonk ML, Cole JC, Watson P, Gillet V, Willett P. SuperStar: Improved knowledge-based interaction field for protein binding sites. J Mol Biol 2001;307:841–859.
23. Nissink JWM, Verdonk ML, Klebe G. Simple knowledge-based descriptors to predict protein–ligand interactions: Methodology and validation. J Comput Aided Mol Des 2000;14:787–803.
24. Chakravarty S, Bhinge A, Varadarajan R. A procedure for detection and quantitation of cavity volumes in proteins. J Biol Chem 2002;30:31345–31353.
25. Yamashita MM, Wesson L, Eisenman G, Eisenberg D. Where metal ions bind in proteins. Proc Natl Acad Sci U S A 1990;87:5648–5652.
26. Hellinga HW, Richards FM. Construction of new ligand binding sites in proteins of known structure: I. Computer-aided modelling of sites with pre-defined geometry. J Mol Biol 1991;222:763–785.
27. Yang W, Lee H-W, Hellinga H, Yang JJ. Structural analysis, identification, and design of calcium-binding sites in proteins. Proteins 2002;47:344–356.
28. Walles AC, Borkakoti N, Thornton JM. TESS: A geometric hashing algorithm for deriving 3D coordinate templates for searching databases: Application to enzyme active sites. Protein Sci 1997;6:2308–2323.
29. Taroni C, Jones S, Thornton JM. Analysis and prediction of carbohydrate binding sites. Protein Eng 2000;13:89–98.
30. Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M. Automated analysis of interatomic contacts in proteins. Bioinformatics 1999;15:327–332.
31. Sobolev V, Edelman M. Modeling the quinone-B binding site of the photosystem-II reaction center using notions of complementarity and contact surface between atoms. Proteins 1995;21:214–225.
32. Sobolev V, Vriend G, Wade RC, Edelman M. Molecular docking using surface complementarity. Proteins 1996;25:120–129.
33. Thompson JD, Higgins DG, Gibson TJ. Clustal-W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
34. Hung LW, Wang IX, Nikaido K, Liu PQ, Ames GF, Kim SH. Crystal structure of the ATP-binding subunit of an ABC transporter. Nature 1998;396:703–707.
35. Burley SK, Petsko GA. Amino-aromatic interactions in proteins. FEBS Lett 1986;203:139–143.
36. Levitt M, Perutz MF. Aromatic rings act as hydrogen-bond acceptors. J Mol Biol 1988;201:751–754.
37. Basharov MA, Volkenstein MV, Golovanov IB, Nauchitel VV, Sobolev VM. The fragment–fragment interaction method: Part I. Estimating molecule–molecule interactions. J Gen Chem USSR 1989;59:435–447.
38. Mitchell JBO, Nandi CL, McDonald IK, Thornton JM, Price SL. Amino/aromatic interactions in proteins—is the evidence stacked against against hydrogen-bonding? J Mol Biol 1994;239:315–331.
39. Felder C., Jiang HL., Zhu WL, Chen KX, Silman I, Botti SA, Sussman JL. Quantum/classical mechanical comparison of cation–pi interactions between tetramethylammonium and benzene. J Phys Chem A 2002;105:1326–1333.
40. Derewenda ZS, Lee L, Derewenda U. The occurrence C-H…O hydrogen-bonds in proteins. J Mol Biol 1995;252:248–262.
41. Bella J, Berman HM. Crystallographic evidence for $C_{\alpha}H\ldots O{=}C$ hydrogen bonds in a collagen triple helix. J Mol Biol 1996;264:734–742.

42. Wahl MC, Sundaralingam M. C-H…O hydrogen bonding in biology. Trends Biol Sci 1997;22:97–102.
43. Chakrabarti P, Chakrabarty S. C-H…O hydrogen bond involving proline residues in α-helices. J Mol Biol 1998;284:867–873.
44. Ghosh A, Bansal M. C-H…O hydrogen bonds in minor groove of A-tracts in DNA double helices. J Mol Biol 1999;294:1149–1158.
45. Luisi B, Orozco M, Sponer J, Luque FJ, Shakked Z. On the potential role of the amino nitrogen atom as a hydrogen bond acceptor in macromolecules. J Mol Biol 1998;279:1123–1136.
46. Leibowitz N, Nussinov R, Wolfson HJ. MUSTA—a general, efficient, automated method for multiple structure alignment and detection of common motifs: Application to proteins. J Comput Biol 2001;8:93–121.
47. Knöfel T, Sträter N. Mechanism of hydrolysis of phosphate esters by the dimetal center of 5′-nucleotidase based on crystal structures. J Mol Biol 2001;309:239–254.
48. Obmolova G, Teplyakov A, Bonander N, Eisenstein E, Howard AJ, Gilliland GL. Crystal structure of dephospho-coenzyme A kinase from *Haemophilus influenzae*. J Struct Biol 2001;136:119–125.
49. Babor M, Sobolev V, Edelman M. Conserved positions for ribose recognition: Importance of water bridging interactions among ATP-, ADP- and FAD-protein complexes. J Mol Biol 2002;323:523–532.
50. Denesyuk AI, Denessiouk KA, Korpela T, Johnson MS. Functional attributes of the phosphate group binding cup of pyridoxal phosphate-dependent enzymes. J Mol Biol 2002;316:155–172.