Published online 3 June 2009

Nucleic Acids Research, 2009, Vol. 37, No. 14 ***4613–4620***
doi:10.1093/nar/gkp488

Cavities in protein–DNA and protein–RNA interfaces

Shrihari Sonavane and Pinak Chakrabarti*

Department of Biochemistry and Bioinformatics Centre, Bose Institute, P-1/12 CIT Scheme VIIM, Calcutta 700 054, India

Received March 12, 2009; Accepted May 19, 2009

ABSTRACT

An analysis of cavities present in protein–DNA and protein–RNA complexes is presented. In terms of the number of cavities and their total volume, the interfaces formed in these complexes are akin to those in transient protein–protein heterocomplexes. With homodimeric proteins protein–DNA interfaces may contain cavities involving both the protein subunits and DNA, and these are more than twice as large as cavities involving a single protein subunit and DNA. A parameter, cavity index, measuring the degree of surface complementarity, indicates that the packing of atoms in protein–protein/DNA/RNA is very similar, but it is about two times less efficient in the permanent interfaces formed between subunits in homodimers. As within the tertiary structure and protein–protein interfaces, protein–DNA interfaces have a higher inclination to be lined by β-sheet residues; from the DNA side, base atoms, in particular those in minor grooves, have a higher tendency to be located in cavities. The larger cavities tend to be less spherical and solvated. A small fraction of water molecules are found to mediate hydrogen-bond interactions with both the components, suggesting their primary role is to fill in the void left due to the local non-complementary nature of the surface patches.

INTRODUCTION

Cavities are defects in proteins (1,2), the interior of which have tightly packed atoms (3–5). Often water molecules occupy these cavities (6,7) and can compensate for the destabilization of reduced hydrophobic and van der Waals interactions (8). Similarly, imperfection in surface complementarity during the complexation between the protein subunits may lead to the location of cavities in the interface (9). Recently the cavities in protein interiors and protein–protein interfaces have been placed in the same general footing in terms of their number, volume, the nature of the cavity-lining atoms/residues and the associated secondary structural features, solvation, etc. (10). From these perspectives we analyze cavities in protein–DNA and protein–RNA interfaces in this work. Though there have been studies aimed at deciphering physicochemical features of these interfaces (11–27), and comparison have been made vis-à-vis those in protein–protein interfaces (28), no detailed study has been undertaken to understand features of cavities formed in protein nucleic-acid interactions (29).

MATERIALS AND METHODS

Datasets

Atomic coordinates of the protein–DNA and protein–RNA complexes were obtained from the Protein Data Bank (PDB) (30). Out of 128 protein–DNA complexes used in (18) and 50 protein–RNA complexes in (27), we retained those determined at least to a resolution of 3 Å to create datasets of 115 and 42 cases, respectively for this analysis. In two PDB files (1du3 and 1k78) the macromolecular assembly consisted of two different monomers interacting with DNA in spatially distinct regions—these were split into two separate protein–DNA complexes, but involving the same DNA. Both monomeric and homodimeric proteins have been considered (59 and 56 cases, respectively) in protein–DNA complexes. However, due to the paucity of data only the monomeric proteins were included in protein–RNA complexes. The atoms that lose at least 0.1 $Å^2$ of the accessible surface area (ASA) in the complex structure as compared to that in the isolated subunit were considered as interface atoms (31,32). In PDB files *O*-phosphotyrosine and selenomethionine atoms are listed under the HETATM records. To avoid locating spurious cavities we considered these atoms as part of protein coordinates (rather than hetero atoms). Also cavities lined by residues with missing atoms were excluded.

Identification and classification of cavities

To be compatible with our earlier analysis (10) cavities were identified using the CASTp (Computed Atlas of Surface Topography of proteins) server (33) located at

*To whom correspondence should be addressed. Tel: +91 33 2355 0256; Fax: +91 33 2355 3886; Email: pinak@boseinst.ernet.in; pinak_chak@yahoo.co.in

http://sts.bioengr.uic.edu/castp/, with the default probe radius of 1.4 Å. Cavity classes considered in this analysis are designated as (i) PD (in the interface formed by a protein subunit and DNA), (ii) PDP (located between DNA and both the subunits of homodimeric proteins) and (iii) PR (in monomeric protein–RNA interface). The first two are illustrated in Figure 1. To be considered as an interface cavity it should have at least 20% of the cavity-lining atoms from both DNA and the protein component. For homodimeric proteins if both the subunits contribute to the cavity it is identified as PDP. Only the cavities with volume $>11.5\,Å^3$ (the volume of the probe with radius 1.4 Å) were included in the analysis. For the identification of water molecules structures determined to a resolution of 2.4 Å or better were used (25) and the PD and PR cavities were classified as solvated or empty based on the presence or the absence of crystallographically determined water molecules in them. Hydrogen bonds involving the water molecule (to protein atoms, as well as to other water molecules in the cavity) were determined using HBPLUS (34). The molecular diagrams were made using MSMS (35) and VMD (36).

Major groove and minor groove atoms

Atomic labels of the base components considered as belonging to the major groove are C5, C6, C8, N6, N7 (for adenine), C5, C6, C8, O6, N7 (guanine), C4, C5, C6, N4 (cytosine), C4, C5, C6, C5M and O4 (thymine). Atomic labels of minor groove atoms are C2, C4, N3, N9 (adenine), C2, C4, N2, N3, N9 (guanine), C2, O2, N1 (cytosine and thymine). Atoms N1 (adenine and guanine), N3 (cytosine and thymine) are not considered in either groove.

Propensity

The propensity of a residue to be a part of a cavity is given as lnP, where

$$P = \frac{(Nx/\sum Nx)}{(Na/\sum Na)}.$$

Nx is the number of atoms of residue type X lining the cavities and $\sum Nx$ is its total number in the interfaces; Na and $\sum Na$ are the corresponding numbers considering all the residue types together. This method is based on counting the atoms, rather than residues, as explained in (10). Likewise, the propensity was also calculated for the occurrence of secondary structural elements (helix, strand and the rest, termed as 'Others') and major/minor groove atoms lining the cavities. Secondary structure assignment was made using the program DSSP (37).

Cavity shape

Rvs was used to ascertain if a cavity was spherical. It is the ratio of volume to surface of a cavity relative to that for a sphere with the same volume.

$$\text{Rvs} = \frac{(\text{Volume/Surface area})_{\text{cavity}}}{(\text{Volume/Surface area})_{\text{sphere}}}$$

Cavity index

The cavity index for a protein–nucleic acid or protein–protein interface was calculated as,

$$\text{Cavity Index}(Å) = \frac{\text{Total volume of interface cavities } (Å^3)}{\text{Average interface area } (Å^2)}.$$

The term in the denominator is the total interface area (31,32) buried between the two components of the complex divided by 2.

RESULTS

For 59 protein–DNA complexes involving monomeric proteins a total of 991 cavities were detected, out of which 149 belonged to the protein–DNA interface and are termed PD cavities (Supplementary Table S1). For 56 structures where the protein component is homodimeric there are 1208 cavities, of which 229 are located in the interface. However, when a contiguous stretch of DNA is in contact with both the subunits (and there are 44 structures), two types of interface cavities are possible, PD involving a single subunit of the protein and the DNA, and PDP that encompasses DNA and both the protein subunits (Figure 1). Only 28 among 229 are PDP cavities (which are found in 15 structures only). Of the remaining 201 interface cavities, because of symmetry we considered only 113 involving only one protein subunit as of type PD. This makes a total 262 PD cavities. Protein–RNA complexes had 971 cavities in total, of which 108 are PR. Approximately 20% of both types of interfaces are devoid of detectable cavities. The detailed information on cavities in individual PDB entries is provided in Supplementary Table S2.

Number and total volume of cavities in interfaces

The average number of PD and PR cavities in protein–nucleic-acid interface is about the same as observed for PP_C in transient protein–protein complexes, and these

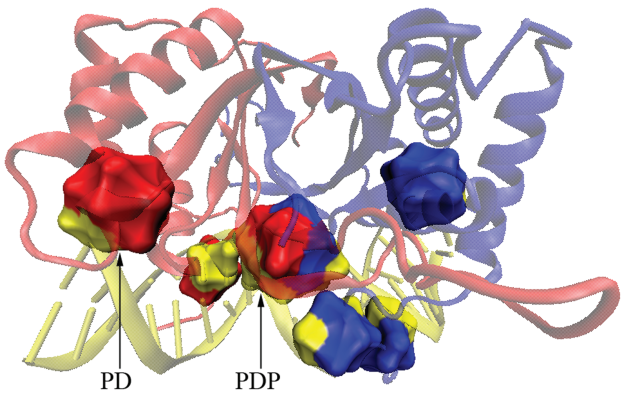

Figure 1. Surface representation of the cavities present at protein–DNA interface. Two types of cavities possible at the interface involving a homodimeric protein are shown using the structure of EBNA-1 Nuclear protein–DNA complex (PDB file, 1b3t); protein and DNA chains are displayed in cartoon.

Table 1. Average values of the total number of cavities and the total cavity volume in different interfaces and protein tertiary structure

	PD	PDP	PR	PP_C	PP_H	Ter_str
Number of cavities	2.3 (2.1)	1.9 (1.2)	2.6 (2.6)	2.4 (2.1)	5.0 (4.5)	15.5 (13.9)
Normalized number[a]	20 (15)	16 (11)	19 (17)	24 (17)	25 (14)	15 (6)
Total cavity volume (Å^3)	82 (114)	181 (212)	97 (211)	97 (127)	324 (498)	486 (517)
Normalized volume[a]	710 (674)	1505 (2278)	726 (762)	970 (1034)	1585 (1439)	473 (245)

The standard deviations are in parentheses. PD, PDP and PR cavities are defined in the text. PP_C and PP_H cavities belong to protein–protein interfaces in heterocomplexes and homodimers, corresponding to Inter_C and Inter_H cavities, respectively, in ref. (10). Ter_str are cavities belonging to protein tertiary structure. The data for the last three columns are from ref. (10). Using 101 monomeric proteins that bind DNA or RNA we recalculated the values for the parameters given in the table for Ter_str cavities and obtained 17 (17), 14 (6), 511 (632) and 417 (207).
[a]Number or volume (given in the previous row) per 2000 interface atoms. The volume of a protein with 2000 atoms is about 49 000 Å^3. Though the PDP cavities contain atoms from the second protein subunit, for normalization we have assumed that both PDP and PD cavities are located in the same number of interface atoms (between one subunit and DNA)—this way their normalized volumes can be compared directly.

are about a sixth compared to that found in protein tertiary structures (Table 1). PP_H cavities found in the obligate interfaces formed between subunits in homodimeric molecules contain about twice the number of cavities as compared to other interfaces. However, as homodimeric interfaces are twice the size of those in heterocomplexes (32) or protein–DNA complexes (11,18) and also the cavities in tertiary structure are contained in a larger number of atoms, we normalized the number to the value expected for an ensemble of 2000 atoms. Compared to the tertiary structure, protein–nucleic-acid interfaces contain about 1.3 times the number of cavities, but the increase is 1.6 times for both the protein–protein interfaces. Considering the normalized volume, relative to the tertiary structure the increase observed in protein–nucleic-acid interfaces is ~1.5 times, whereas for protein–protein complexes the values are 2.1 and 3.4 times for PP_C and PP_H cavities, respectively. Thus the cavities in protein–protein interfaces are larger (especially for homodimers) as compared to those in protein–nucleic-acid interfaces. When present, the PDP cavities occur to almost the same extent as PD cavities, but are more than two times as large.

We also studied if the features of PD cavities differ based on the function of the protein–DNA complexes (18). Results presented in Supplementary Table S3a indicate that the interfaces belonging to excision and/or repair class have the highest number and volume of cavities, intermediate between what is seen in the two types of protein–protein interfaces. Cavities belonging to the enzyme class are similar to PP_C cavities, while those belonging to transcription factors and 'Others' resemble Ter_str cavities. We also looked at the features of PD cavities depending on whether the DNA is single-stranded, double-stranded, or cleaved (Supplementary Table S3b)—the values observed for the last category seem to be slightly on the higher side.

Distribution of cavity volume and shape

The total volume of the cavities in individual interface is poorly correlated with the interface size (defined by number of atoms present in the interface) (Supplementary Figure S1), as was observed in protein–protein interfaces. However, for individual cavities one can use power law or linear fit (Supplementary Table S4) to express the variation of the volume with the number of atoms or residues lining the cavity (Supplementary Figure S2). Typical of tertiary structure cavities (10), approximately five atoms or four cavity-lining residues are needed to accommodate one water molecule. The distribution of volumes of cavities is shown in Figure 2a. Like the cavities in protein–protein interfaces, protein–nucleic-acid interfaces also contain higher percentage of cavities that are larger than

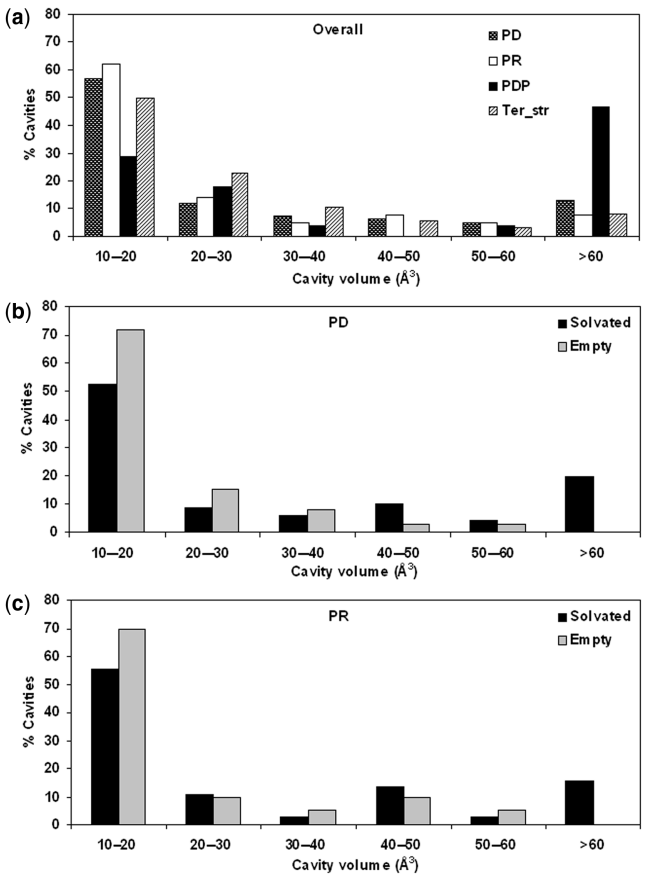

Figure 2. Histogram of cavity volumes. Different cavity types are shown in (**a**). In (**b**) and (**c**) the distribution is shown for solvated and empty cavities (only PD cavities are used for homodimeric proteins); Ter_str cavities are defined in Table 1.

100 Å^3—7.2 and 6.5% for PD and PR cavities, relative to 2.6% in cavities in the tertiary structure. However, the PDP cavities tend to be by far the largest and 46% of them are larger than 60 Å^3. The largest cavity observed in a protein–nucleic-acid interface is of type PDP and shown in Figure 3. The larger cavities are usually solvated (Figure 2b and c). Totally 75% and 66% of PD and PR cavities are solvated, respectively (considering the volume, the percentages are 88% and 85%, respectively). As was inferred from the data in Supplementary Table S3a, when grouped in different functional classes the PD cavities belonging to excision and/or repair enzymes contain more number of larger cavities (>100 Å^3) followed by enzymes, 'Others' and transcription factors (Supplementary Figure S3).

The parameter Rvs—the surface:volume ratio of a cavity as compared to that for a sphere having the same volume as the cavity—indicates how spherical a cavity is; a perfect sphere would have a value of 1.0, with a lower value indicating deviation from a spherical shape. The distribution of Rvs (Supplementary Figure S4) indicates that ~75% PD and PR cavities have value >0.90, whereas 54% of PDP cavities have values <0.90. That the larger cavities—and PDP cavities are mostly large—tend to be of irregular shape can be seen from the histogram of Rvs values for cavities with volume >100 Å^3, having a peak near 0.75.

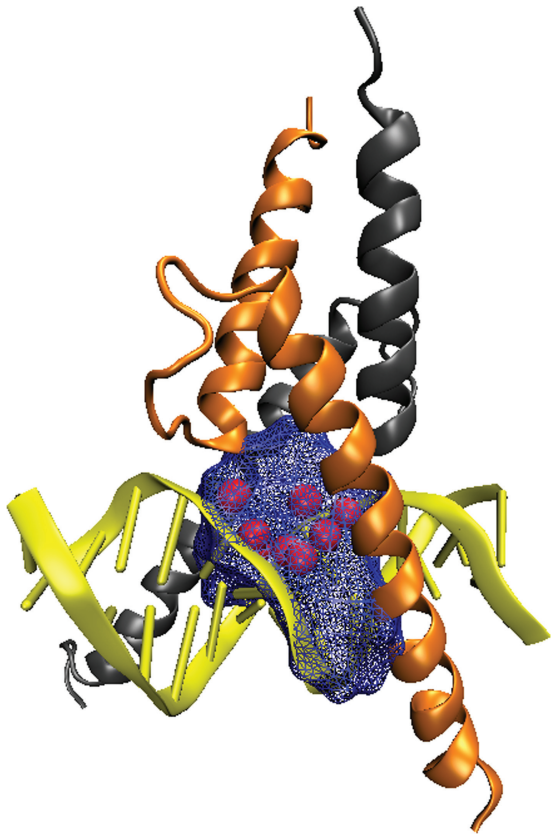

Figure 3. The largest PDP cavity located in 1mdy (volume 828 Å^3, Rvs 0.59 and 9 water molecules).

Preferences of amino acids and nucleotides to be located in cavities

The propensities of amino acids to line the cavities are shown in Figure 4—a large, positive (or negative) value indicates preference (or avoidance), and a value close to zero suggest neutral behavior. Features in PD cavities are quite distinct—hydrophobic residue are favored, hydrophilic residues (and Cys) disfavored. PR cavities show some differences. Among hydrophobic residues, Val, Phe, Leu and Ile are preferred, along with hydroxyl-containing groups (Ser, Thr and Tyr); positively-charged residues (Lys, Arg and His), and Pro and Gly, in particular are disfavored. In general, the avoidance of charged residues and the preference for hydrophobic residues have also been noted for cavities in tertiary structures and protein–protein complexes (10).

Propensities of base, sugar and phosphate moieties to be associated with the cavities (Figure 5) show that bases are favored, whereas phosphate (in particular for PR) is disfavored. If we differentiate the base atoms in protein–DNA interfaces into those belonging to major and minor grooves, it is observed that the latter are favored. Supplementary Figure S7b shows an example of protein–DNA interface with four cavities for which the contribution of base atoms are mostly from minor groove.

Secondary structure preferences

The propensities of different secondary structural elements to be associated with cavities are shown in Figure 6.

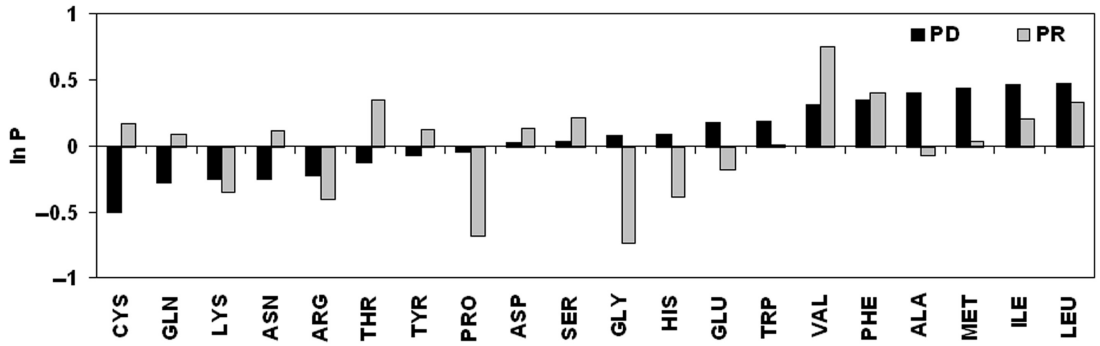

Figure 4. Propensities of residues to be associated with PD and PR cavities.

The significance of the propensity values has been confirmed from the *z*-values (38), shown in Supplementary Figure S6a. In PD cavities, strands are the most preferred element, followed by helices, as has been observed in protein–protein interfaces (10). However, in PR cavities, the helices are the only preferred element. The involvement of strand residues in PD cavities can be seen in Supplementary Figure S7a.

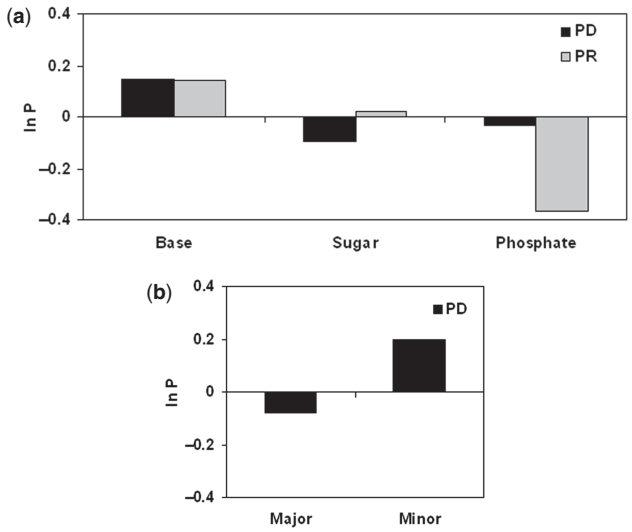

Figure 5. Propensities of nucleotide (**a**) components and (**b**) groove atoms to be associated with cavities. Percentage compositions used in the calculation of propensities for groove atoms are available in Supplementary Figure S5a and b.

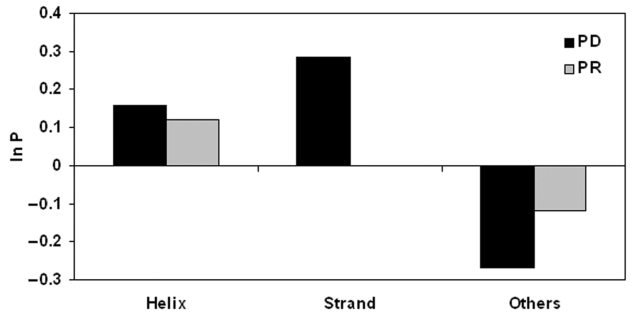

Figure 6. Propensity of cavity-lining atoms to occur in different secondary structural elements. Percentage compositions used in the calculation of propensities are available in Supplementary Figure S5c and d.

Cavity index

The cavity index (described in 'Materials and methods' section) is a measure for interface complementarity. The smaller the cavity index, the more complementary the interface surfaces are. The plot for the distribution of the parameter (Figure 7) indicates a higher average value (0.15) and thus lesser surface complementarity of interfaces formed between the subunits in homodimeric proteins as compared to those in protein–DNA/RNA/protein heterocomplexes. A larger percentage (8.2%) of homodimeric proteins have value >0.45, reflecting the larger size of cavities in these interfaces. Analyzing the protein–DNA complexes in different functional classes we find that those involved in excision and/or repair have lesser interface surface complementarity, having an average value of 0.12 (as compared to 0.09 for enzymes and 0.05 for transcription factors).

Water molecules in cavities and their interactions

If the solvent molecules are disordered (which may happen if there is no strong hydrogen bond interactions holding them, and if the volume available is much larger than what is needed to accommodate them) and/or the resolution of diffraction pattern is not high enough, these are not likely to be seen X-ray crystallographic analyses (10). However, within this limit of structural studies, one finds that 75% of PD cavities contain water molecules—larger than 61–66% observed in PR and PP_C cavities and 50% observed for cavities in tertiary structure (10). A larger proportion of cavities in protein–DNA interfaces were also observed to be filled with water than in the protein interior (29), though the exact proportions were different from the values given here.

Though water molecules are generally believed to mediate interaction between protein and DNA, the number of such molecules was found to be only 6%, with the majority (76%) being involved either to solvate the protein or the DNA atoms at the interface (39). From our analysis of water molecules in the interface cavities (Table 2) we observe that water molecules that have

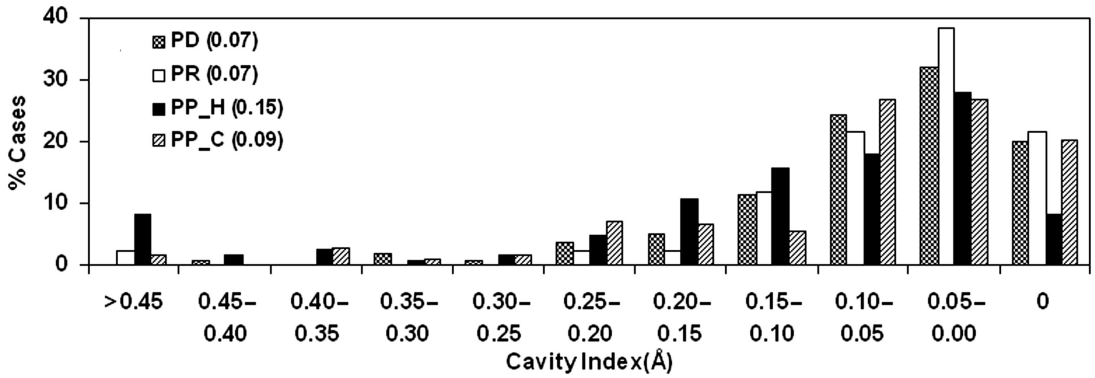

Figure 7. Distribution of cavity index in different types of interfaces. Cavity indices for individual protein–protein complexes are provided in Supplementary Table S6.

Table 2. Water molecules in PD and PR cavities and their hydrogen-bonding pattern

	PD	PR
Number of solvated cavities	119	38
Number of water molecules	259	73
Number with HB (both)	95 (231)[a]	12 (55)[a]
HB (single)	143	53
HB (none)	21	8

The labels—both, single and none—indicate the number of water molecules that are hydrogen bonded to both the subunits across the interface, or just to one, or neither of them.
[a]The numbers in parentheses are the water molecules having contacts within 4.0 Å (not necessarily hydrogen bonds) from both the sides.

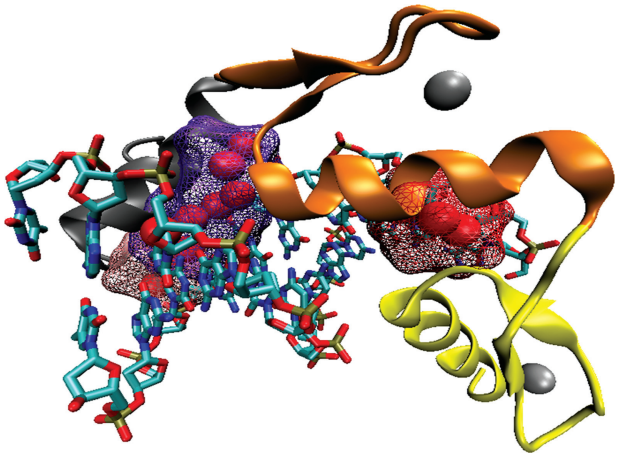

Figure 8. Three interface cavities in the structure of a zinc finger protein (PDB file, 1aay), with three domains, assigned using SCOP (43), shown in distinct colors. The cavity in red has a volume of 92 Å^3 and contains three water molecules, in blue (47 and 2), and in violet (180 and 7). Out of 12 interface water molecules only one forms hydrogen bonds with protein and DNA both (bridging water), 9 forms HBs with only one component (protein or DNA) and remaining two do not form any HB with either component. If we consider contacts (instead of HB), 11 out of 12 are within 4 Å from both the sides.

direct hydrogen bonds to both protein and DNA are just 37%, while 55% water molecules form hydrogen bonds to one component only. For protein–RNA complexes the corresponding figures are 16% and 73%, respectively. An example of water molecules in cavities is shown in Figure 8. Though the water molecules considered here are a subset of those used in the earlier study, the general conclusion that the majority of water molecules are not involved in bridging the two sides with hydrogen bonding still holds true. (However, if we consider the water molecules that are in contact with both the sides, though not necessarily forming hydrogen bonds, 90% of them would satisfy the condition). About 10% water molecules just fill in the cavities without forming any specific hydrogen bond to either side. This observation for protein–DNA complexes matches with what was found in homodimeric interfaces, though the interfaces formed in protein–protein heterocomplexes showed a higher percentage (~50%) of bridging solvent molecules (10).

For the bridging water molecules one can ask the question if they can buffer the unfavorable electrostatic interaction between the negatively-charged phosphate group and a carboxylate side chain, or close positioning of pairs of hydrogen-bond acceptors or donors at the interface. Results in Supplementary Table S5 indicate that the extent of occurrence of water between phosphate and a negative or a positive residue is in the ratio of about 1:2, suggesting that the solvent molecules are present more to fill in the void left in the interface, rather than to neutralize the like charges coming close to each other. Water molecules are found to occur in the ratio of ~1:6 between two acceptors (or donors) and between a donor and an acceptor.

DISCUSSION

Interfaces—general features of cavities and atomic packing

Overall protein–nucleic-acid interfaces resemble those in protein–protein heterocomplexes; however, there can be variation between different functional classes of proteins interacting with DNA. Data presented in Supplementary Table S3a and Supplementary Figure S3 indicate that the category of excision and/or repair enzymes has the highest number and volume of cavities in the interface and transcription factor the least. The former tend to be intermediate between the values for the homodimers and protein–protein heterocomplexes and the normalized values for the latter are more like the cavities in tertiary structures (Table 1).

The atomic packing, as measured by gap volume index, indicated a poorer packing of the protein–RNA complexes as compared to the ones involving DNA (19). However, when the quality of packing is evaluated by measuring the fraction of buried atoms at the interface the former seems to be better packed (25). The cavity index indicates the total volume of cavities present per unit size of the interface—based on its value (Figure 7), or the normalized number (or volume) of cavities (Table 1), there does not seem to be much difference between the interfaces formed by a protein with another protein, DNA or RNA; if at all, the protein–nucleic-acid interfaces appear to be slightly more tightly packed. Because a greater number of usually larger cavities are found in the obligate interfaces formed between the subunits in homodimers (10), the quality of packing as indicated by any of the three features considered here is the poorest for homodimers.

It is worth commenting on the gap volume index (40) used by Jones *et al.* for protein–DNA/RNA interfaces (12,19) and cavity index used here. The former is the ratio of the volume available between the solvent accessible surfaces of the two components of the complex, divided by the interface area. No distinction is made between the interior of the interface and the periphery, where the shape complementarity is likely to be poorer with water molecules getting in between the two interacting surfaces (41). Consequently, the values of the gap volume index tend to be larger, 3.3(±1.8) and 2.6(±0.87) for protein–RNA and protein–dsDNA complexes,

respectively (19). The cavity index, though based on a similar ratio, is restricted to volume of the cavities located inside the interface, should be indicative of the packing in the more important interface interior. Nadassy *et al.* (29) have observed poor correlation between gap volume index and other parameters delineating shape complementarity at interface and suggested that the former may be a more proper representation of the surface complementarity at the periphery of the interface.

Interface cavities in multi-subunit/domain proteins

Because of the involvement of different numbers of protein subunits the interface cavities in homodimeric proteins were segregated into PD and PDP classes (Figure 1). But it turned out that the division was apt based on physical features also. When present, PDP cavities are larger than the PD ones (Table 1), as can be seen in the structure of leucine zipper (Figure 3). The presence of larger cavities at the interface between DNA and the two subunits of homodimeric proteins is also likely to be found in higher oligomeric proteins and also at the domain–domain boundary of multi-domain proteins. Indeed, in the case of zinc-finger protein, one can see that the cavities located between domains 1 and 2, and domains 2 and 3 are much larger than those present between the individual domains and DNA (Figure 8). Using gap volume index the monomeric proteins were found to have more tightly packed protein–DNA interfaces than dimeric proteins (12)—the possibility of the occurrence of larger PDP cavities involving dimeric proteins may be the reason for this.

Higher preference of minor groove atoms for cavities

Protein–DNA interactions may entail a large conformational change in the DNA molecule (12,13,15). The minor groove atoms (Supplementary Figure S8) are found to be involved to a greater extent relative to the major groove atoms in the interface cavities (Figures 5b and Supplementary Figure S6b). As found in protein–protein interfaces and in protein tertiary structures, of all the secondary structural elements the β-sheets have a higher inclination to be involved in interface cavities in protein–DNA complexes (Supplementary Figure S7a). Only in protein–RNA interfaces the helices contribute more to cavities (Figure 6). Helices are known to be disfavored in the RNA recognition sites (20,26), and here we find that these elements in the protein–RNA interfaces are also more likely to contain cavities, though the structural reason behind this is not quite obvious.

Interface and cavity water molecules

The interfaces in protein–protein heterocomplexes contain ~10 water molecules per 1000 $Å^2$, the number being seven for protein–DNA interfaces (28). As a typical interface contains one atom per 9.9 $Å^2$ (41), the above numbers of water molecules can be assumed to belong to 100 interface atoms. Considering only the solvated cavities if we find out the number of water molecules per 100 cavity-lining atoms, we obtain values of 18.0 and 20.7 for protein–protein and protein–DNA complexes, respectively. Thus ordered water molecules can be identified more in solvated cavities than the overall interface. However, if we consider the number of water molecules per 100 polar atoms (all atom types excluding C), we obtain values of 41.5 (protein–protein) and 44.0 water molecules, which are rather close to the value of 37 obtained for protein–protein complexes (42). Both the solvated and empty cavities in protein–DNA interfaces are composed of 49% polar atoms, but in protein–protein complexes the contribution of polar atoms in these cavities are 45 and 29%, respectively, indicating a role of polar atoms in immobilization of solvent molecules.

As in protein–protein interfaces (10), some of the cavities contain ligands, though the number is very meager. Only seven interfaces have been found to contain ligands, usually ions, along with water molecules (Supplementary Table S7).

CONCLUSIONS

The packing of atoms in interfaces can be judged by the occurrence of cavities—the normalized number (and volume) of cavities and cavity index provide a perspective different from the ones commonly used (28) to judge the quality of interface formed by a protein with another protein chain, DNA or RNA. Using these features the interfaces resulting from the transient interactions between macromolecules seem to be very similar. Except for protein–RNA interfaces in which helical residues have a higher propensity to harbor cavities, β-sheet residues are more prominent in the cavities in other interfaces, as well as in tertiary structures. Likewise, minor grooves in DNA are propitious for the location of cavities. Majority of the water molecules located in cavities are hydrogen bonded to one of the components only. Inter-domain or inter-subunit space is likely to be associated with larger cavities in protein–DNA complexes. With the availability of more structures it may be possible to translate such information on the size and nature of cavities to quantitative binding energetics.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Department of Science and Technology; Department of Biotechnology fellowship (to S.S.).

Conflict of interest statement. None declared.

REFERENCES

1. Connolly,M.L. (1986) Atomic size packing defects in proteins. *Int. J. Pept. Protein Res.*, **28**, 360–363.
2. Hubbard,S.J., Gross,K.H. and Argos,P. (1994) Intramolecular cavities in globular proteins. *Prot. Eng.*, **7**, 613–626.
3. Richards,F.M. (1974) The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.*, **82**, 1–14.
4. Richards,F.M. (1977) Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.*, **6**, 151–176.

5. Tsai,J., Taylor,R., Chothia,C. and Gerstein,M. (1999) The packing density in proteins: standard radii and volumes. *J. Mol. Biol.*, **290**, 253–266.
6. Rashin,A.A., Iofin,M. and Honig,B. (1986) Internal cavities and buried waters in globular proteins. *Biochemistry*, **25**, 3619–3625.
7. Williams,M.A., Goodfellow,J.M. and Thornton,J.M. (1994) Buried waters and internal cavities in monomeric proteins. *Protein Sci.*, **3**, 1224–1235.
8. Takano,K., Yamagata,Y. and Yutani,K. (2003) Buried water molecules contribute to the conformational stability of a protein. *Protein Eng.*, **16**, 5–9.
9. Hubbard,S.J. and Argos,P. (1994) Cavities and packing at protein interfaces. *Protein Sci.*, **3**, 2194–2206.
10. Sonavane,S. and Chakrabarti,P. (2008) Cavities and atomic packing in protein structures and interfaces. *PLoS Comput. Biol.*, **4**, e1000188.
11. Nadassy,K., Wodak,S.J. and Janin,J. (1999) Structural features of protein-nucleic acid recognition sites. *Biochemistry*, **38**, 1999–2017.
12. Jones,S., van Heyningen,P., Berman,H.M. and Thornton,J.M. (1999) Protein-DNA interactions: a structural analysis. *J. Mol. Biol.*, **287**, 877–896.
13. Tolstorukov,M.Y., Jernigan,R.L. and Zhurkin,V.B. (2004) Protein-DNA hydrophobic recognition in the minor groove is facilitated by sugar switching. *J. Mol. Biol.*, **337**, 65–76.
14. Lejeune,D., Delsaux,N., Charloteaux,B., Thomas,A. and Brasseur,R. (2005) Protein-nucleic acid recognition: statistical analysis of atomic interactions and influence of DNA structure. *Proteins*, **61**, 258–271.
15. Sarai,A. and Kono,H. (2005) Protein-DNA recognition patterns and predictions. *Annu. Rev. Biophys. Biomol. Struct.*, **34**, 379–398.
16. Sathyapriya,R., Vijayabaskar,M.S. and Saraswathi,V. (2008) Insights into protein-DNA interactions through structure network analysis. *PLoS Comput. Biol.*, **4**, e1000170.
17. Paillard,G. and Lavery,R. (2004) Analyzing protein-DNA recognition mechanisms. *Structure*, **12**, 113–122.
18. Biswas,S., Guharoy,M. and Chakrabarti,P. (2009) Dissection, residue conservation, and structural classification of protein-DNA interfaces. *Proteins*, **74**, 643–654.
19. Jones,S., Daley,D.T., Luscombe,N.M., Berman,H.M. and Thornton,J.M. (2001) Protein-RNA interactions: a structural analysis. *Nucleic Acids Res.*, **29**, 943–954.
20. Treger,M. and Westhof,E. (2001) Statistical analysis of atomic contacts at RNA-protein interfaces. *J. Mol. Recognit.*, **14**, 199–214.
21. Jeong,E., Kim,H., Lee,S.W. and Han,K. (2003) Discovering the interaction propensities of amino acids and nucleotides from protein-RNA complexes. *Mol. Cells*, **16**, 161–167.
22. Cusack,S. (1999) RNA-protein complexes. *Curr. Opin. Struct. Biol.*, **9**, 66–73.
23. Allers,J. and Shamoo,Y. (2001) Structure-based analysis of protein-RNA interactions using the program ENTANGLE. *J. Mol. Biol.*, **311**, 75–86.
24. Morozova,N., Allers,J., Myers,J. and Shamoo,Y. (2006) Protein-RNA interactions: exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. *Bioinformatics*, **22**, 2746–2752.
25. Bahadur,R.P., Zacharias,M. and Janin,J. (2008) Dissecting protein-RNA recognition sites. *Nucleic Acids Res.*, **36**, 2705–2716.
26. Ellis,J.J., Broom,M. and Jones,S. (2007) Protein-RNA interactions: structural analysis and functional classes. *Proteins*, **66**, 903–911.
27. Biswas,S., Guharoy,M. and Chakrabarti,P. (2008) Structural segments and residue propensities in protein-RNA interfaces: comparison with protein-protein and protein-DNA complexes. *Bioinformation*, **2**, 422–427.
28. Janin,J., Bahadur,R.P. and Chakrabarti,P. (2008) Protein-protein interaction and quaternary structure. *Q. Rev. Biophys.*, **41**, 133–180.
29. Nadassy,K., Tomas-Oliveira,I., Alberts,I., Janin,J. and Wodak,S.J. (2001) Standard atomic volumes in double-stranded DNA and packing in protein–DNA interfaces. *Nucleic Acids Res.*, **29**, 3362–3376.
30. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
31. Chakrabarti,P. and Janin,J. (2002) Dissecting protein-protein recognition sites. *Proteins*, **47**, 334–343.
32. Bahadur,R.P., Chakrabarti,P., Rodier,F. and Janin,J. (2003) Dissecting subunit interfaces in homodimeric proteins. *Proteins*, **53**, 708–719.
33. Binkowski,T.A., Naghibzadeh,S. and Liang,J. (2003) CASTp: computed atlas of surface topography of proteins. *Nucleic Acids Res.*, **31**, 3352–3355.
34. McDonald,I.K. and Thornton,J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, **238**, 777–793.
35. Sanner,M.F., Olson,A.J. and Spehner,J.C. (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, **38**, 305–320.
36. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph*, **14**, 33–38, 27–38.
37. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
38. Karpen,M.E., de Haseth,P.L. and Neet,K.E. (1992) Differences in the amino acid distributions of 3(10)-helices and alpha-helices. *Protein Sci.*, **1**, 1333–1342.
39. Reddy,C.K., Das,A. and Jayaram,B. (2001) Do water molecules mediate protein-DNA recognition? *J. Mol. Biol.*, **314**, 619–632.
40. Laskowski,R.A. (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph*, **13**, 323–330, 307–308.
41. Bahadur,R.P., Chakrabarti,P., Rodier,F. and Janin,J. (2004) A dissection of specific and non-specific protein-protein interfaces. *J. Mol. Biol.*, **336**, 943–955.
42. Rodier,F., Bahadur,R.P., Chakrabarti,P. and Janin,J. (2005) Hydration of protein-protein interfaces. *Proteins*, **60**, 36–45.
43. Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.