conference papers

An overview on 2-methyl-2,4-pentanediol in crystallization and in crystals of biological macromolecules

Kanchan Anand, Debnath Pal[*] and Rolf Hilgenfeld

Department of Structural Biology and Crystallography, Institute of Molecular Biotechnology, Beutenbergstr. 11, D-07745 Jena, Germany. E-mail: dpal@imb-jena.de

2-Methyl-2,4-pentanediol (MPD) is the most popular chemical additive used for crystallization of biological macromolecules. However, the mechanism of its action on proteins in aqueous solution is not well understood. We have carried out a systematic analysis of the conformation and environment of MPD molecules bound to proteins. We find that the majority of MPD molecules adopt their most stable conformer. They prefer to bind to hydrophobic sites with a distinct preference for leucine side chains. Most MPD binding sites involve amino-acid residues in helical or β-sheet structures. MPD binding to proteins is penetrative, leading to displacement of water molecules in grooves and cavities (sometimes ligand-binding and active sites) on the protein surface. This results in a large reduction of solvent-accessible area, which can have significant implications for protein stability. The packing of the MPD molecules by the protein is not optimal and usually some other solvent molecules are also bound along with it. Our analysis suggests that MPD is not as strong a denaturant as often suggested. It promotes stabilization of the protein by preferential hydration, which is facilitated by attachment of MPD molecules to the hydrophobic surface.

Keywords: 2-methyl-2,4-pentanediol; crystallization additive; surface area; conformation; interactions

1. Introduction

Although collectively polyethylene glycols (PEGs) of different molecular weights are more popular, 2-methyl-2,4-pentanediol (MPD) is the single most successful agent promoting crystallization of biological macromolecules (as found in the website http://wwwbmcd.nist.gov:8080/bmcd/bmcd.html). This small 'polyalcohol' has properties midway between PEG and organic solvents (McPherson, 1985). It has dual lipophilic and hydrophilic, *i.e.* amphiphilic, character, which makes it suitable for binding to highly heterogeneous protein matrices. Its short hydrocarbon chain is made up of single bonds, allowing it to flexibly adapt to the shape of and efficiently interact with depressions on the protein surface. Incidentally, the dielectric constant of pure MPD is 25, which means that when added to aqueous solutions, it will lead to a reduction of the dielectic constant. Analyses of aqueous MPD/protein systems have shown that increasing concentrations of MPD have a decreasing effect on the dielectric constant of the medium (Arakawa & Timasheff, 1985). In crystallization experiments, MPD can act as precipitant through a combination of activities, including competition for water, hydrophobic exclusion of protein solutes, lowering of the solution dielectric, and detergent-like effects (McPherson, 1998). Its high solubility in water over a wide range of temperatures is an additional advantage, especially in low-temperature experiments. In addition, it can also be used as a cryoprotecting agent (Schneider, 1996), making it extremely invaluable for the protein crystallographer. MPD indeed appears to be an ideal chemical additive for crystallization.

Our current interest in MPD as a crystallization agent has been propelled by the successful use of MPD in our recent crystallization of the first coronaviral main proteinase (TGEV M[pro]) and its location in the electron density maps at 1.96 Å resolution (Anand *et al.*, 2002). Additionally, the usefulness of MPD has been proven beyond doubt by the successful crystallization of the gigantic ribosomal complexes (7-15% v/v MPD) (Thygesen *et al.*, 1996). An average of around 20% (v/v) MPD is used in protein crystallization experiments where the reagent is applied as a precipitant, although this varies over a wide range, from 0.5 to 82% (URL: http://www.cstl.nist.gov/div831/carb/gilliland_group/database/database.html) (Gilliland & Ladner, 1996). Encouraged by this fact, we attempt here to systematically analyse MPD-mediated crystallization results, and determine structural aspects of its interaction with proteins. Although there are several studies (Banumathi *et al.*, 2001; Steiner *et al.*, 2001; Weiss *et al.*, 2000) elaborating on MPD under various crystallization conditions for a given protein, a comprehensive study on proteins in general is yet to be done. The results give insight as to how MPD prefers to interact with the protein surface, thereby promoting formation of good crystals. This, we hope, will help promote a more rational use of MPD as an universal agent in crystallization experiments.

2. Materials and methods

2.1. Creating the database

The words "MPD" or "PENTANEDIOL" were searched for in the Protein Data Bank (PDB) files using the UNIX "egrep" command at our local mirror of the PDB (http://www.rcsb.org/pdb/) (Berman *et al.*, 2000), to create a selected list of 781 coordinate files which contained 2353 polypeptide and 603 oliogonucleotide chains (separate or in complex with protein). From this dataset (Dataset 1), we eliminated the redundant protein files by evaluating the pair-wise sequence similarity of the polypeptide chains using the CLUSTALW program (Thompson *et al.*, 2000). A dendrogram was drawn to visually inspect the clustering (threshold value: 90% identity) of the best-aligned sequences and the file containing the best structural model (indicated by best resolution and R-factor) was picked up from each of the clusters. We neglected oligonucleotide chains and focussed on polypeptides with less than 90% sequence identity (Dataset 2, 377 polypeptide chains). Of all the PDB files selected at this step, only 77 actually contained coordinates for MPD molecule(s) (Dataset 3). If a coordinate set contained multiple copies of MPD molecules as a consequence of non-crystallographic symmetry, we selected the unique molecule(s) by giving preference to the MPD(s) with the lowest average B-factor. In the end, the MPD molecules were checked for consistency of steric contacts. Those that showed a large number of short contacts were eliminated from the database, yielding 80 remaining unique MPD molecules (Dataset 4, 66 PDB files).

MPD exists in two enantiomeric forms, *R* and *S*, due to a chiral centre at the C4 atom (Fig. 1). The absolute majority of the C4 atoms in the database had definite stereochemistry, which was verified by measuring its deviation from the plane defined by the C3-O4-C5 atoms. This yielded values of 0.486(±0.054) Å and -0.488(±0.072) Å for *R* and *S* enantiomers, respectively. We ended up with 36 *R* and 44 *S* structures in our database, with an average temperature factor of 36(±19) Å^2 for all MPD molecules. 82% of the MPD atoms in the database had an occupancy of 1.00 and 99% had an occupancy of at least 0.50. We have systematically checked for consistency and uniformity of formatting and labelling of MPD molecules. We followed a convention of writing the MPD atom coordinates in the order C1-C2-O2-CM-C3-C4-O4-C5 (Fig. 1). Any MPD molecule written in any other labelling convention was converted to the above

order. We found that the C1 and CM positions were most frequently exchanged. The rectified MPD coordinate files are available for public use in the user location http://www.imb-jena.de/www_sbx/debnath/ mpd/mdp_coord.html.

Average trends representative of all polypeptides in the PDB were calculated from a nonredundant protein database (1164 files) compiled previously for another study (Brandl *et al.*, 2001).

2.2. Calculation of geometric and interaction parameters involving MPD

The geometrical parameters of MPD were calculated on Dataset 4. Other interaction parameters were also calculated from this database. The nonbonding contact partners of MPD were evaluated for protein and heteroatom molecules using a threshold distance of 3.8 Å. When an atom from a protein residue or a heteroatom molecule had more than one contact to the MPD molecule, the one with the shortest distance was selected. If more than one atom from a given residue made contacts with the MPD, we counted the residue only once while calculating residue-based statistics. The secondary structure of the proteins was assigned using the PROCHECK suite (Laskowski *et al.*, 1993). The accessible surface areas (ASA) of the individual MPD/protein pairs, in complexed and uncomplexed states, were calculated using the program NACCESS (Hubbard, 1992), which is an implementation of the Lee and Richards (1971) algorithm. The probe size for the surface accessibility calculation was kept at 1.4 Å. The relative value of accessibility of the residue X was evaluated as a percentage by comparing against a standard tripeptide Ala-X-Ala in extended conformation.
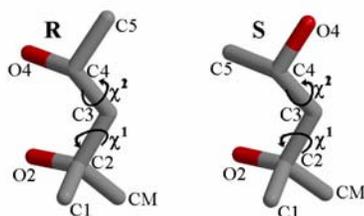
Figure 1 Representation of the MPD molecule in *R* and *S* enantiomeric forms. The sp^3-hybridized C4 atom is chiral. The atoms are labelled according to the convention used in this study (for details, see text). The torsion angles χ^1 and χ^2 are defined by atoms C1-C2-C3-C4 and C2-C3-C4-C5, respectively.

3. Results and discussion

Our database of proteins that had been crystallized in the presence of MPD contained 781 PDB files. Among these, 377 polypeptide chains displayed < 90% pairwise sequence identity and were considered an unbiased data set. The average length of the polypeptide chains was 250 residues, with the shortest being 13 and the longest 1015. A subset of only 77 PDB files contained MPD coordinates, of which 66 were with MPDs devoid of serious errors. If two or more MPD molecules were related by noncrystallographic symmetry, only one was selected, yielding a total of 80 unique MPD molecules. The three-dimensional structure analysis below is based on the 36 *R*- and 44 *S*- enantiomeric MPD molecules constituting this data set.

3.1. Geometrical parameters of bound MPD molecules

The average bond angles and bond distances of MPD molecules are given in Table 1. Comparing the variability of the bond lengths and angles from the standard deviation data suggests the isomers to be somewhat pliable. The difference of 2-4° in the mean values for the C1-C2-C3 / CM-C2-O2 / C3-C4-O4 angles between *R*- and *S*-

MPD is rather large. The length of the C1-C2 bond and the bond angles around C2 are especially deviant. To find out if such differences are artifacts, we compared the respective parameters with five small-molecule complexes containing MPD (1 in *R*, 2 each in *S* and racemate form) stored in the Cambridge Structural Database (http://www.ccdc.cam.ac.uk) (Allen & Kennard, 1993; Refcodes: KOFPAW, BACXIM10, NIRQIO, NOSVOG, TECYIJ). In these cocrystals with different molecules, the C1-C2 bond distance varied by as much as 0.055 Å, while the CM-C2 bond distance differed by only 0.010 Å. There were similar large differences in the bond lengths between C2-C3, C3-C4, and C4-C5. These were also reflected in the bond angle values with very large differences among equivalent (protein MPD / 'small molecule' MPD) bond angles. It is rather surprising to find such digressions and we checked the R-factor and the coordinate ESD values to ensure the reliability of our observation. The worst R-factor was 0.123 for TECYIJ, while the best was 0.065 for KOFPAW. The maximum errors in the fractional coordinates in all the structures were only in the third significant digit. Due to the large data-to-parameter ratio for small-molecule structures, there is usually no bias in the refined final model. On the contrary, heavy stereochemical restraints are the norm to produce correct geometry during structure refinement of macromolecules (Kuriyan *et al.*, 1986). However, if there are persistent trends, then despite such restraints, appreciable standard deviation values in the calculated statistical parameters are consistently reproduced. It appears from the small-molecule MPD structures that the single-bond distances are quite variable, possibly due to large inductive effects. These putative environment-induced effects probably arise from different strengths of nonbonding interactions modulated by varying effective partial charges on the atoms. This may result in contextual differences in binding energetics for different MPD molecules. If such effects are present during docking onto proteins, it can have profound implications on the stabilization and hydration of the protein surface. As a note of caution, however, we must emphasize that the postulation made on environment-induced effects would be more convincing if we had more small-molecule structures to compare our statistical findings.

3.2. Conformation of MPD molecules

The distribution of conformations of MPD molecules bound to proteins can be seen from Figure 2. The expected values of torsion angle combinations (χ^1,χ^2) are $(300^\circ,180^\circ)$ for *R* and $(180^\circ, 180^\circ)$ for *S*, respectively, because the relative position of O4 and the C5 atoms are interchanged when the MPD molecule exists in *R*- or *S*-form (Fig. 1). These combinations are preferred to avoid unfavourable 1-5 contacts, and bring the O2 and O4 atoms to the nearest proximity, thus allowing the formation of an intramolecular hydrogen bond. With any other combination of torsion angles, there is at least one additional O…C or C…C contact below 3.0 Å (see legend to Fig. 2). The steric repulsion due to such unfavourable short contacts clearly drives the (χ^1,χ^2) combinations to only one preferred value supported by the intramolecular hydrogen bond; this trend is largely reflected in the 'small-molecule structures' as well. Interestingly, Weiss *et al.* (2000) performed a similar analysis, and their results show the cluster at $(300^\circ,180^\circ)$ to be less populated than that at $(60^\circ, 180^\circ)$. This is exactly opposite in our case. We surmise that this difference could be a consequence of inconsistent atom labelling since the distribution of data points in the other regions of the plot is essentially similar. It also appears that the assignment of MPD atom types in the PDB has been done quite consistently while interpreting the electron density maps; this is apparent from the distribution of the conformers (Fig. 2), whose distributions are consistent with their noncovalent potential energies. It is also evident from Figure 2 that the most stable conformation involving the intramolecular hydrogen

bond in MPD is preserved in the majority of its interactions with proteins, and only sometimes interactions are strong enough to drive the MPD away from the lowest energy state. Another interesting consequence of this is that the polar oxygen and nonpolar carbon atoms are disposed in such a manner that the molecule is uniformly divided into hydrophilic and hydrophobic halves.

The average B-factor of R- and S-MPD molecules in the database is 33($\pm$17) Å^2 and 40($\pm$21) Å^2, respectively. The relatively large difference in the mean values for the R and S isomer is un-explained. The (χ^1, χ^2) distribution (Fig. 2) does not indicate an uneven distribution that could indicate that S-MPD might be in a less favorable conformation connected with higher mobility.
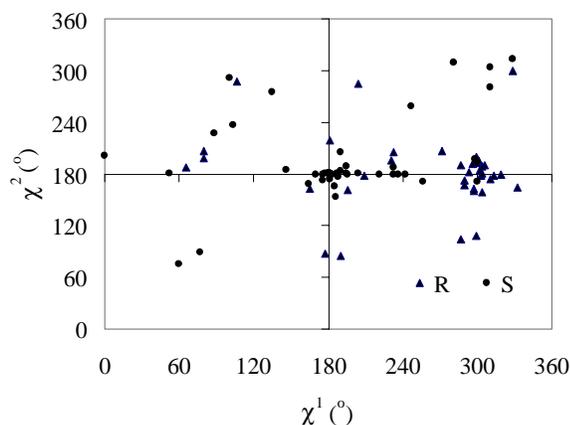
Figure 2 Torsion angle distribution (χ^1, χ^2) of MPD molecules. The nomenclature is described in Figure 1. The 1-5 C...C short contacts (≤ 3.0Å) debar the following (χ^1, χ^2) combinations in R-MPD: (CM-C5 | 60º,60º), (C1-C5 | 60º,300º) , (C1-C5 | 300º,60º); similar short contacts in S-MPD are: (CM-C5, C1-C5 | 60º,60º), (CM-C5 | 180º,300º), (C1-C5 300º,60º). Bad O...C steric contacts (≤ 3.0Å) for R-MPD are: (O4-C1 | 60º,180º), (O4-CM | 60º,300º), (O2-C5 | 180º, 60º), (O4-CM | 180º,180º), (O2-C5, O4-C1 | 300º,300º) and for S-MPD: (O4-C1 | 60º,180º), (O2-C5, O4-CM | 180º,60º), (O4-C1 | 300º, 180º), (O2-C1 | 300º,300º).

Table 1 Mean values and the associated standard deviations (in parentheses) of the bond distances and bond angles of the MPD molecules found in PDB coordinate sets

Bond	Distance (Å)		Bond	Angle(º)	
	R	S		R	S
C1-C2	1.52(3)	1.51(5)	C1-C2-CM	113(8)	114(7)
CM-C2	1.52(2)	1.52(2)	C1-C2-O2	106(4)	106(5)
O2-C2	1.45(3)	1.45(3)	C1-C2-C3	111(3)	109(4)
C2-C3	1.53(2)	1.54(3)	CM-C2-O2	108(5)	106(3)
C3-C4	1.52(3)	1.51(3)	CM-C2-C3	109(3)	112(3)
C4-C5	1.53(3)	1.52(3)	O2-C2-C3	110(3)	110(3)
C4-O4	1.43(3)	1.43(2)	C2-C3-C4	116(4)	116(3)
			C3-C4-O4	112(5)	108(4)
			C3-C4-C5	109(4)	110(4)
			C5-C4-O4	108(3)	108(3)

3.3. MPD binding to proteins

73% of the 80 MPD molecules in our database interact with just one protein subunit in the crystal. An overwhelming 86% of the R-MPD molecules are in this category while this fraction is much lower at 64% for the S isomer. There have been earlier reports of the existence of a limited number of sites on the protein surface that attract many different organic molecules, regardless of their sizes and polarities (Mattos & Ringe, 1996). Active sites of enzymes qualify in this category and we looked into this aspect by analyzing the "SITE" card in the PDB and found that 20% of the MPD

molecules are located near or at the active site. All these MPD molecules, with one exception, had contacts to only one protein subunit. In general, only about 28% of the MPD molecules in our database made contacts with more than one protein subunit. A similar fraction made symmetry-related contacts.

3.3.1. Environment overview

Residue-based interaction statistics offer a broad overview on the role of MPD as a valuable chemical additive. We have checked the distribution over the amino acid sequence of the residues binding MPD. Only a negligible fraction of binding sites exists where the residues come from stretches that are local in sequence. A histogram of the distribution in the amino acid sequences of residues making up the MPD binding site shows a Gaussian with the centre at 100 residues in between (data not shown). This indicates that MPD anchors non-local regions of the protein which possibly helps in lowering mechanical fluctuations on the surface (loss of conformational entropy). We also checked if the number of MPD atoms contacting the protein has any correlation to the number of residues it anchors at the binding site. We find that the number of MPD atoms in contact increases linearly with the number of anchored residues (correlation coefficient of 0.81). The results show that the bulk of the MPD prefers to bind to atoms of a few residues; if this would not be the case, the correlation we have obtained here would be exponential. The number of anchored residues can also be correlated with the solvent-accessible area (ASA); with increasing number of anchored residues, the MPD ASA should decrease. This is what we find and an exponential trend is expected for optimal packing. We get a general trend from our data, but the non-exponential character of the plot (data not shown) indicates that the protein atoms cannot pack the MPD in an optimal way.

A large number of water molecules were found to be in contact with the O2 and O4 hydroxyl groups of MPD. The average B-factor of the water molecules contacting MPD is 30($\pm$14) Å^2 (value for all water molecules 35($\pm$14) Å^2). The water mostly satisfies the unsaturated hydrogen-bonding potential of the MPD molecules.

3.3.2. Residue preferences

Figure 3 shows a clear preference for Leu as the residue most preferred in the MPD binding site. When taken together for both R- and S-MPD, this is almost double that of the next favoured residue, Tyr. This indicates that MPD prefers to bind to surface-exposed (Fig. 3 (inset)) hydrophobic patches, and the binding is probably driven by hydrophobic interactions. Our results show a 3:2 preference for non-polar over polar protein atoms for MPD-protein contacts.

3.3.3. Binding surface

An entropically driven binding entails release of solvent water molecules attached to the protein surface undergoing burial and a concomitant decrease in conformational entropy of the protein residues constituting the surface. The eventual change in degrees of freedom is an important entropic component in the free energy of binding of the MPD. An approximate estimate of this change can be obtained from the analysis and quantification of the accessibilities of the surfaces undergoing burial or exposure on binding. The average solvent-accessible surface of an unbound MPD is 277($\pm$3) Å^2. When bound to the protein, the solvent accessibility changes to 75 Å^2 on average, with a large standard deviation of 53 Å^2. This indicates a predominantly buried state of the molecule with only around one fourth of its surface exposed, precluding attachment of the MPD to shallow depressions on the protein surface. For such non-occlusive binding, the solvent-accessible surface area would have an average of around 140 Å^2, *i.e.*, half the solvent-accessible surface area of the unbound molecule. We also quantified the interface area buried upon

MPD binding by calculating the solvent accessibility of the MPD-complexed protein, and the protein and MPD separately. We found that binding of one MPD molecule on average buries an interface area of 320($\pm$53) Å^2, which is much higher than the average accessible surface area of the MPD molecule itself. This is the sum of the surface of the protein and MPD buried on binding and it means that on average, MPD binding into a groove reduces the area accessible on both binding partners to bulk solvent by around 320-(277-75) $\approx$120 Å^2 [*Average surface area reduced on binding = (Average of total surface area of protein + MPD buried on binding) – (Average of total accessible surface area of unbound MPD – Average of total accessible surface area of MPD on binding)*]. Using Eisenberg and McLachlan's (1986) equation for empirically estimating free energies of solvation, $\Delta G_s = \Sigma$ ASP * ΔASA (where ASP = atomic solvation parameter and ΔASA=accessible surface area buried), one can roughly estimate the gain in free energy for a hydrophobic surface burial. From our estimates of large effective burial of surface area, even with large errors in the ΔASA determination and low ASP values, ΔG_s is expected to be significant. This is because the free energy of the native protein is itself very small and the numbers of bound MPDs are expected to be in multiples (assuming that some MPDs are undetectable by X-ray crystallography, but nevertheless bound). Thus, the binding of MPD can have significant bearing on the thermodynamic stability of the protein.
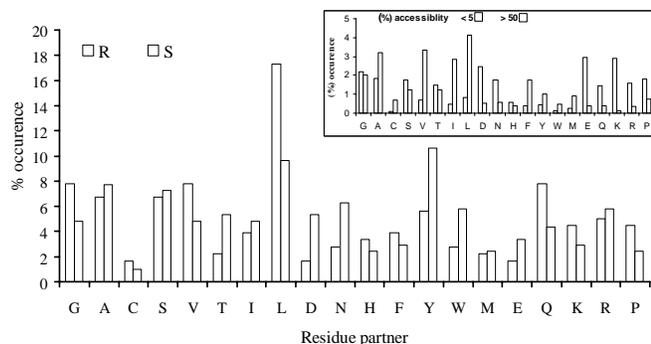
Figure 3 Frequency of the distribution of residues (denoted by one-letter code) contributing atom(s) to the binding site of MPD. In the inset is the general distribution of the relative surface accessibility of amino-acid residues in proteins.

3.3.4. Secondary structures binding MPD

Helices. Around 35% of the MPD molecules bound to proteins are associated with residues forming α or 3_{10} helices (Fig. 4). When bound to helices, the preference is mainly for residues at the centre and not near the termini of the helix. We looked into the preferences of amino acids in helices that bind MPD. Leu is the most prominent residue followed by Arg and Lys. In Arg and Lys, the side-chain guanidinium and amino group, respectively, are the preferred binding partners for MPD. The occurrence of leucine is significantly larger than for the rest of the amino acids. It is known that Leu has a high propensity to be in α-helical conformation and therefore the high frequency of Leu–MPD contacts is not unexpected. However, most surface-exposed helices have a periodic distribution of hydrophobic and hydrophilic residues and usually the buried half of the helix is hydrophobic and the exposed hydrophilic. Our results suggest that unfavourably exposed leucine residues are acquiring an MPD molecule to get buried from solvent. This is exemplified in Figure 5A by the MPD complex of coenzyme F420-dependent tetrahydromethanopterin reductase (Shima *et al.*, 2000). The *R*-MPD

molecule is in its most favoured (300°,180°) conformation, as a result of which the molecule has its surface uniformly divided into polar and nonpolar faces. The nonpolar side of the molecule is bound to a casket created by the intersection of two α-helices from the same subunit of the protein. The C1 MPD atom buries a part of the Ala2180 residue, while the CM atom sits on top of the Pro2155 pyrrolidine. The C3 MPD atom covers the C$^{\delta 1}$ atom of the solvent-exposed Leu, and similarly C5 covers the C$^{\gamma 2}$ atom of Ile2170 from a neighbouring strand. The MPD molecule appears to be clamped by the guanidinium group of Arg2238. Therefore, apart from covering the exposed hydrophobic surface, here the MPD also effectively restricts the Arg side chain to a single rotameric state, which can have a beneficial effect for crystallization. Interestingly, the protein residues lining the binding pocket are highly conserved (Shima *et al.*, 2000). The solvent accessibility of the MPD is only 32 Å^2. We checked if MPD binding sites are always formed by more than one helix, as in this example. We found that in a majority of the cases, multiple helices formed the binding site, suggesting that helix junctions could be a potential receptacle for the MPD molecules.

β-Sheets. Around 32% of residues that bind MPD come from β sheets (Fig. 4). Since β sheets are present in various topologies in proteins, it is sometimes difficult to quantify if a bound MPD molecule is on the edge or face of the sheet. We circumvent this by investigating if the binding is preferred with the edge strands or the middle strand(s) of the sheet. Using the program DSSP (Kabsch & Sander, 1983), we denominated all MPD-binding strands as "edge" or "middle" and found a majority (two-thirds) to be at the edge. Of these, an overwhelming fraction (> 90%) are antiparallel strands.
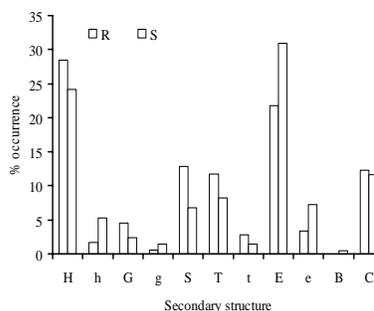
Figure 4 Distribution of the residues in various secondary structural elements that bind MPD molecules. The secondary structure denominations (Laskowski *et al.*, 1993) are H for α-helix core, h: α-helix termini, G: 3_{10}-helix core, g: 3_{10} helix termini, S: non-hydrogen-bonded bend, T: hydrogen-bonded turn (middle residues), t: hydrogen-bonded turn (terminal residues), E: β-strand core, e: β–strand termini, B: β-ladder, C: irregular secondary structure.

Since antiparallel β strands, on average, tend to be less hydrophobic than parallel ones (Richardson, 1981), the question arises if MPD binding is also facile to less hydrophobic surfaces. This is important in light of an earlier study suggesting that interactions between the peptide group and polyols are unfavourable (Gekko, 1981). For an insight into this, we looked into the preference for main-chain contacts with MPD from edge strands and found an appreciable fraction of cases. Among these, we found an interesting case where the MPD molecule is on the outer surface of the barrel where it is predominantly interacting through hydrogen bonds with the protein (Fig. 5B); hydrophobic interactions apparently do not play a major role here. The C1 atom of MPD is covering the Leu C$^{\delta 1}$ and the CM is stacked on top of the Val95 - Gly96 peptide bond, but does not protect the exposed thiol group of Cys37 from solvent exposure. The O2 is nearest to the Gly96 C$^{\alpha}$ atom and likely involved in a water-mediated hydrogen bond to the Gly96 carbonyl oxygen (Fig. 5B). The Ser98 O$^{\gamma}$ strongly hydrogen-bonds to the O4 atom and the Gly35 main chain is stacked against C5 of the MPD. The solvent-

accessible area of the MPD is only 23 Å^2. This indicates that pockets with apparent lower hydrophobicity can also be potential binding sites for MPD. Nevertheless, the essential requirement of hydrophobicity for facilitating MPD binding is reflected from the general preference of aromatic residues (> 25%). Ser and Thr are also found in appreciable numbers, similar to Lys and Arg, as in MPD-binding sites involving helices, indicating that while the nonpolar surface of the MPD attaches itself to its complementary protein surface, the hydroxyl groups can provide additional interactions. This mode of attachment may help in keeping the protein residues in anchored positions, with a beneficial role for crystallization.

3.3.5. Binding of MPD to substrate-binding sites

Penchant of organic molecules for binding to a limited number of sites, like ligand-binding and active sites, have been studied (Mattos & Ringe, 1996). We looked into the few cases of MPD binding to active sites in our database and found that Ser is present, in large numbers, followed by Thr, and together these two make up The structure of the coronavirus main proteinase, TGEV M^{pro}, almost 25% of the residues that bind MPD at substrate-binding sites. an example (Fig. 5C). An MPD molecule binds near the S2 and S3 recently determined in our laboratory (Anand et al., 2002), provides subsites in the substrate-binding site, between the two β-barrel core domains. This MPD molecule is also interacting with a long loop (residues 184 to 199) connecting domains II and III; this loop is involved in substrate binding. The C1 of the MPD contacts the C^β of Leu164, O2 forms a hydrogen bond with the main-chain oxygen of Thr47 and C5 stacks with the peptide bond connecting Asp186 and Gln187. The protons attached to C4 and O4 can interact with the imidazole π electrons of His41. The physicochemical natures of the substrate recognition sites of many proteinases are exhaustively explored, and although they are not always hydrophobic, they have hydrophobic subsite(s), which can be potential binding site(s) for MPD molecule(s).
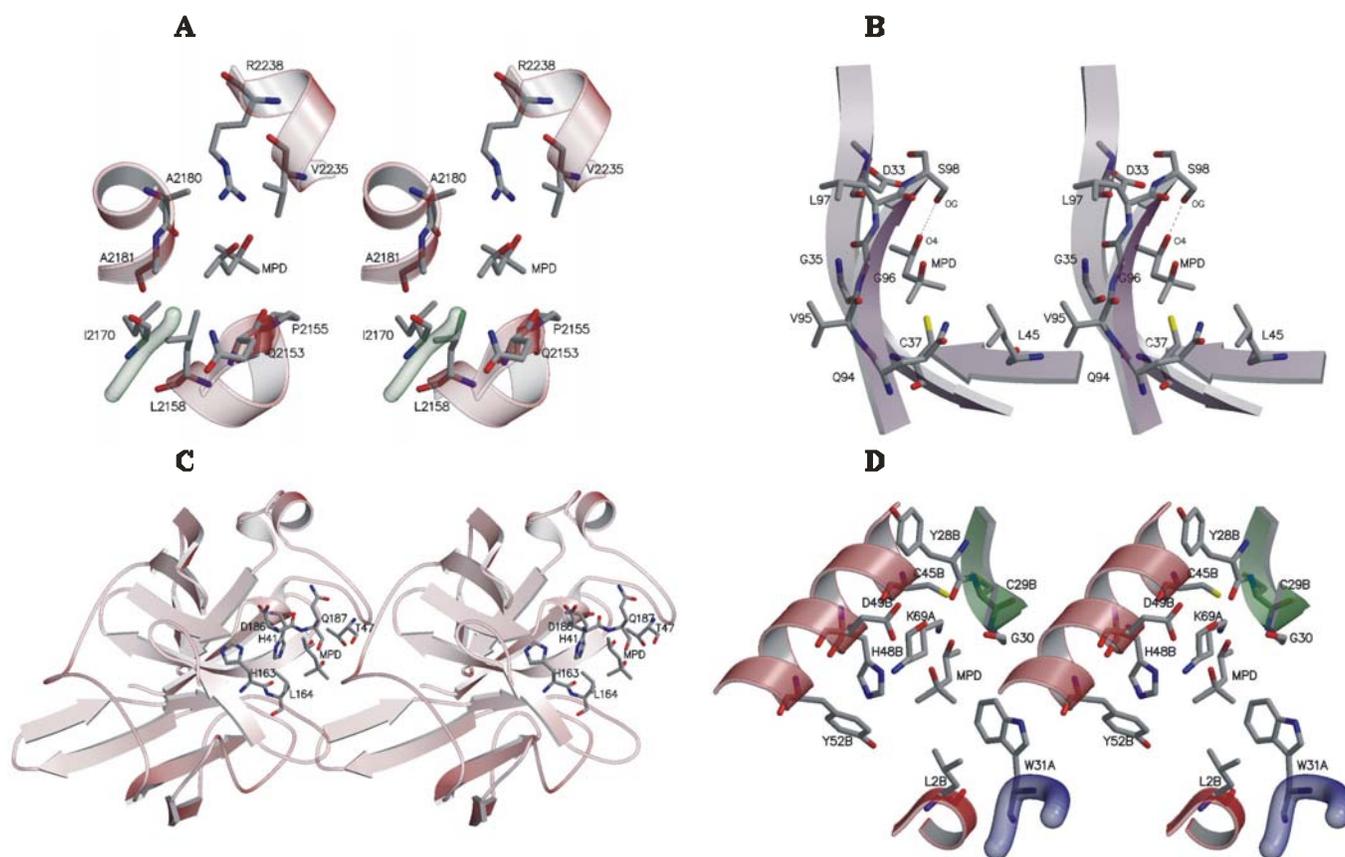
Figure 5 Examples, in stereo, of MPD molecules bound to the protein matrix. All protein residues within 3.8 Å from the MPD molecule are drawn. Residue types are in one-letter code and the sequence numbers are as given in the PDB files (along with the subunit name, if there is more than one). Oxygen atoms are shown in red, nitrogen in blue, sulfur in yellow and carbon in grey. (A). An R-MPD molecule bound to a site created by α-helical residues from a single protein subunit. The coordinates were taken from PDB file 1F07, coenzyme F420-dependent tetrahydromethanopterin reductase solved at 2.0 Å resolution (Shima et al., 2000). (B). An R-MPD molecule bound to the concave outer surface of a β sheet present in a barrel topology. The example was taken from PDB file 1NCO, apo-carzinostatin, solved at 1.8 Å (Kim et al., 1993). A single strong hydrogen bond between O4 of MPD and O^γ of Ser98 is indicated. (C). Figure showing the first two domains of TGEV M^{pro} where the R-MPD molecule binds to an active site cleft interacting with two additional residues from a loop connecting the second and the third domain. (D). An S-MPD molecule in association with two protein subunits. The coordinates are taken from PDB file 1JLT, vipoxin complex, solved at 1.4 Å (Bhanumathi et al., 2001). The figure was drawn using Molscript (Kraulis, 1991).

3.3.6. Intersubunit binding of MPD

The solvent-accessibility statistics indicate dominant burial of hydrophobic surface resulting in a substantial contribution of entropy in binding of MPD to proteins. Definitely, the interaction of the geometry is not optimised, because the binding partners are not in the correct orientations [for example, the methyl groups of MPD are all interacting with the aromatic ring edges and not faces, which is

known to be the more stable arrangement due to formation of CH…π hydroxyl groups of MPD cannot contribute as much electrostatic energy as can water, due to a much lower effective partial charge. Nevertheless, maximization of enthalpic gain is expected in thermodynamic equilibrium. This is highlighted in Figure 5D, where the MPD binds to more than one protein subunit. The O2 atom of MPD is making a hydrogen bond with the Lys69A amino group (2.93 Å). The O4 atom is bonded to the amide NH of Gly30B.The CM atom is in a casket created by the side chains of Leu2B, His48B and Tyr52B. The C3 and C5 atom make contacts with Trp31A and Cys45B, respectively. At first sight, it appears that the interaction bonds (Brandl *et al.*, 2001)]. However, the protein side chains themselves are interacting in a network of stabilizing bonds. Some of these are weak interactions: the Leu2B side chain contacts the face of a Trp ring giving rise to CH…π bonds; the side chain of His48B interacts with the Tyr52B aromatic ring in an edge-to-face interaction; the carboxylate group of Asp49B accepts a CH…O bond from a Tyr28B phenyl ring-edge proton. It is remarkable that these interactions are not disturbed by the presence of the non-proteinous MPD. This suggests that the binding interactions of MPD are entropy-driven and any gain in enthalpy of the system is not at the expense of reduction of the enthalpy of the protein itself. This particular MPD was found to have only 20 Å^2 of its surface area exposed to the solvent.

3.3.7. Crystal contacts

Since MPD is mostly buried inside the protein molecule, its chances of making crystal contacts are minimized. Indeed the ratio of crystal-to normal MPD…protein atom-atom contacts is ~1:9. The situation is similar in intersubunit contacts, with most of the MPD molecules (>90%) preferring to bind a major part or all of its surface to a single subunit. There are no significant trends with regard to the types of residues binding to MPD molecules across crystal contacts. Tyr, Leu, Ala and Ser have a higher-than-average number of contacts. This is in accord with the previous observation that there are no particular residue compositions of the surface patches makingcrystal contacts (Carugo & Argos, 1997). In light of our observation above of a preference for hydrophobic residues in MPD binding sites (Fig. 3) and nothing similar with regard to crystal contacts, it can be argued that in most cases MPD does not initiate the crystal contacts, but rather fills up voids on the surface, thereby stabilizing the protein, as well as increasing the number of interactions stabilizing lattice formation. This is further reinforced by a look at the secondary structures of the residues involved in crystal contacts. Around two thirds of the residues making crystal contact(s) involving MPD are located in either helices or   strands, although normally, these elements of secondary structure make crystal contacts less often than turns or loops (Argos, 1988). It is interesting to note that the O4 atom of MPD is engaged in many of the few crystal contacts observed, almost double the number of O2 contacts. This trend is not apparent when we analyse MPD contacts in general, suggesting that the O4 end of the MPD molecule prefers to remain solvent-accessible more often than O2. This is consistent with the hydrophobic binding mode of the MPD; however, given our small database, it is rather speculative to adjudge the O2 end of MPD to be more hydrophobic resulting in any kind of binding-mode preference.

3.4. Denaturing properties of MPD

There have been some previous discussions on the manner MPD-containing solvent interacts with proteins (Kita *et al.*, 1994). MPD strongly lowers the surface tension of water (*i.e.,* the surface free energy of water), reflecting its amphiphilic character (Hammes & Schimmel, 1967; Pittz & Bello, 1971). This makes it surface-active and drives it to seek contact with the nonpolar residues in proteins. This is essentially reflected in our analysis above but at variance with the experimental observations that MPD is preferentially excluded from proteins (Arakawa *et al.*, 1990; Pittz & Timasheff, 1978). It appears that MPD binding at equilibrium is largely concentration-dependent and penetrative burial of MPD molecules at some particular loci (as we have seen for most of the cases in our analysis) is only possible if there is preferential association of MPD with a complementary protein surface. This possibly results in impermeable patches that cannot be further penetrated by water molecules, although they may trap some of them. The consequence is an excess of water at the protein surface that thermodynamically leads to preferential hydration (promoted by covering of nonpolar surface patches). This is achieved by proper juxtaposition of the MPD molecule, i.e. by maximizing the burial of the protein surface (Fig. 5).

During precipitation, the structure of the protein molecules is identical in the two end states of the process; the chemical nature of the contacts between protein and solvent remains largely unchanged unless there is a huge concomitant burial of surface. If such a process occurs under supersaturated conditions, the concentration of MPD is expected to be rather high, making denaturation of the protein an issue. Although the origin of the protein-denaturing character of MPD (Lee & Lee, 1987; Arakawa *et al.*, 1990) has been thoroughly investigated, and an intrinsic preference for interaction with both end states of the unfolding equilibrium is possible, our analysis suggests MPD not to be a forceful denaturing agent. It does not promote diverging of protein charges, whereby the repulsion is weakened and MPD can penetrate to the newly exposed nonpolar residues, interact favourably with them, and, in this manner, stabilize the unfolded structure. In such a case we would expect many of the proteins crystallized using MPD to be in non-native states (which is obviously not the case).

4. Conclusions

The conformation of the MPD during its interaction with proteins is in its most stable state for a majority of cases. In this conformation, the surface of the MPD is uniformly divided into a hydrophobic and a hydrophilic surface. MPD prefers to bind mostly to nonpolar residues such as Leu, Tyr and Val. This is a convincing indicator of the hydrophobic nature of the binding. Additionally, however, substantial cases exist where polypeptide main-chain atoms bind the MPD, in agreement with the amphiphilic character of the compound. It is remarkable that most MPD binding sites are formed by residues in regular secondary structural elements, mainly helices and β sheets. The major sites of binding are the junction of two helices and the edge of β-sheets. We also found that MPD on average binds to five protein residues usually well separated along the polypeptide sequence, and this number is linearly correlated with the interface area it covers on binding. Accessible surface area analysis of MPD shows its preference to bind in grooves and cavities on the protein surface. This allows it to occlude large amounts of interface area from bulk solvent, with major entropic implications for protein stability. The strategy for MPD binding involves optimization of enthalpy of interaction without disturbing the network of interacting protein residues too much. The general character of the molecule is only mildly denaturing and it promotes hydration of proteins via association with hydrophobic protein surface patches. This indicates the general role for MPD as an agent that stabilize the protein by filling up voids and cavities on the protein surface rather than a glue for binding proteins in lattice together. Thereby it indirectly facilitates proper crystallization; suggesting that its use in optimal concentrations can be very beneficial in crystallization and corroborating its usefulness as a valuable chemical additive.

conference papers

References

Allen, F. H. & Kennard, O. (1993). *Chem. Design. Autom. New,s* **8**, 31-37.

Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. & Hilgenfeld, R. (2002). *EMBO J.* **21**, 3213-3224.

Argos, P (1988). *Protein Eng.* **2**, 101-113.

Arakawa, T. & Timasheff, S. N. (1985). *Methods Enzymol.* **114**, 49-77.

Arakawa, T., Bhat, R. & Timasheff, S. N. (1990). *Biochemistry*, **29**, 1924-1931.

Banumathi, S., Rajashankar, K. R., Notzel, C., Aleksiev, B., Singh, T. P., Genov, N., Betzel, C. (2001). *Acta Cryst*. D**57**, 1552-1559.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res*. **28**, 235-242.

Brandl, M., Weiss, M. S., Jabs, A., Sühnel, J. & Hilgenfeld. R. (2001). *J. Mol. Biol.* **307**, 357-377.

Carugo, O. & Argos, P. (1997). *Protein Sci.* **6**, 2261-2263.

Eisenberg, D. & McLachlan, A. D. (1986). *Nature*, **319**, 199-203.

Gekko, K. (1981). *J. Biochem. (Tokyo)*, **90**, 1643-1652.

Gilliland, G. L. & Ladner, J. E. (1996). *Curr. Opin. Struct. Biol.* **6**, 595-603.

Hammes, G. G. & Schimmel, P. R. (1967). *J. Am. Chem. Soc.* **89**, 442-446.

Hubbard, S. (1992). *NACCESS: A program for calculating accessibilities.* http://sjh.bi.umist.ac.uk/naccess.html.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577-2637.

Kim, K. H., Kwon, B. M., Myers, A. G. & Rees, D. C. (1993). *Science*, **262**, 1042-1046.

Kita, Y., Arakawa, T., Tiao, L, -Y. & Timasheff, S.N. (1994). *Biochemistry*, **33**, 15178-15189.

Kraulis, P. (1991). *J. Appl. Cryst.* **24**, 946-950.

Kuriyan, J., Petsko, G. A., Levy, R. M. & Karplus, M. (1986). *J. Mol. Biol.* **190**, 227-254.

Laskowski, R. A., MacArthur M. W., Moss D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283-291.

Lee, J. C. & Lee. L. L. Y. (1987). *Biochemistry*, **26**, 7813-7819.

Lee, B. & Richards, F. M. (1971). *J. Mol. Biol*., **55**, 379-400.

Mattos, C. & Ringe, D. (1996). *Nature Biotechnol.* **14**, 595-599.

McPherson, A. (1985). *Methods Enzymol.* **114**, 120-125.

McPherson, A. (1998). *Crystallization of Biological Macromolecules*, New York: Cold Spring Harbor Laboratory Press.

Pittz, E. P. & Timasheff, S. N. (1971). *Arch. Biochem. Biophys.* **146**, 513-524.

Pittz, E. P. & Timasheff, S. N. (1978). *Biochemistry*, **17**, 615-623.

Richardson, J.S. (1981). *Adv. Protein Chem.* **34**, 167-339.

Schneider, T. R. (1996). Proceedings of the *IIIrd International School and Symposium on Synchrotron Radiation (ISSRNS'96)*, May 31 – June 8, Jaszowiec, Poland.

Shima, S., Warkentin, E., Grabarse, W., Sordel, M., Wicke, M., Thauer, R. K. & Ermler, U. (2000). *J. Mol. Biol.* **300**, 935-950.

Steiner, R.A., Rozeboom, H. J., de Vries, A., Kalk, K. H., Murshudov, G. N., Wilson, K. S. & Dijkstra, B.W. (2001). *Acta Cryst*. D**57**, 516-526.

Thompson, J. D., Plewniak, F., Thierry, J. & Poch, O. (2000). *Nucleic Acids Res.* **28**, 2919-2926.

Thygesen, J., Krumbholz, S., Levin, I., Zaytzev-Bashan, A., Harms, J., Bartels, H., Schlunzen, F., Hansen, H. A. S., Bennett, W. S., Volkmann, N., Agmon, I., Eisenstein, M., Dribin, A., Maltz, E., Sagi,I., Morlang, S., Fua, M., Francheschi, F., Hilmer, R. M., Boddeker, N., Sharon, R., Anagnostopoulous, K., Peretz, M., Geva, M., Berkovitch-Yellin, Z. & Koyama, T. (1996). *J. Crystal Growth*, **168**, 308-323.

Weiss, M. S., Palm, G. J. & Hilgenfeld, R. (2000). *Acta Cryst.* D**56**, 952-958.

Weiss, M. S., Brandl, M., Sühnel. J., Pal, D. & Hilgenfeld, R. (2001). *Trends Biochem. Sci.* **26**, 521-523.