# Systematic analysis of the probabilities of formation of bimolecular hydrogen-bonded ring motifs in organic crystal structures

**Frank H. Allen,*[a] W. D. Samuel Motherwell,[a] Paul R. Raithby,[b] Gregory P. Shields[a,b] and Robin Taylor[a]**

[a] *Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, UK CB2 1EZ. E-mail: allen@ccdc.cam.ac.uk*

[b] *Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK CB2 1EW*

***Received* (*in Cambridge, UK*) *15th September 1998, Accepted 5th November 1998***

A methodology has been developed for characterising hydrogen-bonded ring motifs formed between two organic molecules without any prior knowledge of the topology or chemical constitution of the motifs. The method has been implemented by modifying the current Cambridge Structural Database (CSD) System programs. All intermolecular ring motifs comprising $\leqslant$20 atoms formed with N—H···N, N—H···O, O—H···N and O—H···O hydrogen bonds in organic structures in the CSD have been classified. The 75 bimolecular motifs occurring in $>$12 structures in the CSD are described in terms of their graph sets and chemical functionalities. Motifs are ranked according to their frequency of occurrence and according to their probabilities of formation, *i.e.* their frequency relative to the number of possible motifs which could have formed. These probabilities provide insights into the relative robustness of known and potential supramolecular synthons.

## Introduction

It is well known that the stronger hydrogen-bond motifs found in organic systems[1] can be used to direct the synthesis of supramolecular complexes, *e.g.* in crystal engineering.[2–6] The conceptual relationship between crystal engineering and organic synthesis has led to the term *supramolecular synthon*[5] being proposed for structure-directing motifs involving non-covalent bonds. Many of the synthons identified so far involve N—H···N, N—H···O, O—H···N and O—H···O H-bonds,[5–7] which confer a degree of robustness, hence reproducibility, in supramolecular retrosynthesis, while knowledge of these common patterns is also important in other applications, *e.g.* the modelling of protein–ligand interactions, and in *ab initio* crystal structure prediction. A number of early studies explored the nature of H-bonded motifs in classes of compounds with particular functional groups, *e.g.* carbohydrates,[8] carboxylic acids[9] and amides.[10] Similar motifs had been recognised previously in inorganic systems.[11] Other studies have examined the effect of H-bond cooperativity and resonance-assistance on the robustness (strengths) of H-bonded systems.[12–14]

In order to describe the topology of H-bonded motifs and networks systematically, a graph-set approach has been suggested.[15,16] This provides a description of H-bonding patterns in terms of chains (C), rings (R), finite complexes (D) and intramolecular H-bonds (S). The degree of the pattern (*n*, the number of atoms comprising the pattern), together with the number of donors (d) and the number of acceptors (a), are combined to form the graph-set designator $X^a_d(n)$.[16] This description does not distinguish between patterns of the same degree that have different numbers of bonds in the constituent fragments or different arrangements of the donors and acceptors. However, these purely topological descriptors have proved useful in decoding differences between the packings adopted in polymorphic systems, *e.g.* in L-glutamic acid[17] and iminodiacetic acid.[18] Patterns are distinguished on the basis of their level, the first level describing patterns involving crystallographically equivalent hydrogen bonds (if any), the second level involving two such H-bonds, and similarly for higher levels. Chemically equivalent patterns may become apparent at different levels depending on the presence or otherwise of crystallographic symmetry. The graph set nomenclature provides a basic description of H-bonded synthons and can aid the identification of preferred motifs. Recently a systematic general search for $R^2_2(8)$ motifs has been performed, to explore the chemical diversity of the functional groups which adopt this topology.[19]

The aim of our present work is to perform a computerised analysis of the non-covalent motifs that occur in the $>$160 000 crystal structures in the Cambridge Structural Database (CSD, October 1996).[20] This is a data-driven approach which is more general than any previous analysis, and is designed to establish the topologies, chemical constitutions and numbers of occurrence ($N_{obs}$) of non-covalent motifs in an objective manner. While $N_{obs}$ values are interesting, their interpretation is complicated by the fact that they depend on the number of times that the various donor and acceptor groups occur in the CSD. Thus, the carboxylic acid cyclic dimer motif may have a high $N_{obs}$ simply because carboxylic acids are common in the CSD. To correct for this effect, we determine the probability of occurrence, through which $N_{obs}$ for each motif is related to $N_{poss}$, the number of times the motif could possibly occur.

Since we can assume that the probability of formation of a motif across many crystallographic instances is related to its robustness[5] (*i.e.* how reliably a motif may be exchanged from one network structure to another), then the analysis has the potential to reveal motifs that may act as novel supramolecular synthons in crystal engineering applications. Because of the broad scope of the overall analysis, we have subdivided the work. Thus, our initial investigations have concentrated on the identification of intermolecular cyclic motifs formed between pairs of molecules by medium to strong H-bonds involving N—H and O—H donors and N or O acceptors.

These cyclic patterns encompass many of the supramolecular synthons already identified.[5] In performing this study, we have also extended the computer methodology so as to generate graph-set descriptors for the motifs which were located. A preliminary account of this work, summarising the 24 most significant motifs, has appeared elsewhere.[21]

As with other CSD investigations, there is the possibility of bias in data selection, *i.e.* the CSD is not a random sample of crystal structures and may not provide a representative picture of the typical environments of the fragments comprising the ring motifs. In particular, there is no guarantee that the degree of competition between different motifs will be the same for all motifs studied. For certain motifs used extensively in crystal design, a particular effort may have been made to avoid competing motifs. Alternatively, attempts may have been made to investigate structures where the expected motif did not occur, *e.g.* the carboxylic acid dimer. However, these considerations are likely to apply to a relatively small number of structures, although certain motifs may be particularly affected. For fragments occurring in a large number of diverse molecules represented in the CSD, as in this study, such bias should not be a significant issue.

## Methodology

### Probabilities of motif formation

To assess the probability of formation of a particular motif ($P_m$) across the complete CSD, it is necessary to count the total number of motifs ($N_{obs}$) that actually occur, and then to compute the total number of motifs that could have occurred ($N_{poss}$) in all structures recorded in the CSD (computation of $N_{poss}$ is discussed below). Thus:

$$P_m = N_{obs}/N_{poss} \quad (1)$$

Alternatively, we can compute a structural probability ($P_s$). If $S_{obs}$ is the number of structures that contain a particular motif, and $S_{poss}$ is the number of structures that contain the component functional groups (the donor and acceptor groups that comprise the motif), then:

$$P_s = S_{obs}/S_{poss} \quad (2)$$

The quantity $S_{poss}$ is easier to compute than $N_{poss}$, hence $P_s$ provides a useful check on the computation of $P_m$, since $P_m$ should be equal to $P_s$ if $P_m$ is independent of the number of motifs $N_{poss}$ which could form in a structure. It should be remembered that both $P_s$ and $P_m$ are based solely on the sample of structures represented in the CSD.

$P_{symm}$ is defined as the percentage of chemically symmetric motifs which are also crystallographically symmetric.

The principal steps in deriving the probabilities were to:

(i) Identify potential donors and acceptors in each structure.

(ii) Search for intermolecular H-bonds between the donors and acceptors.

(iii) Find the shortest intramolecular (covalent) paths between donors and acceptors and hence locate intermolecular ring motifs.

(iv) Count ring motifs and classify them in terms of their sizes, topologies and chemical constitutions.

(v) Identify the chemical constitutions of the donor and acceptor fragments comprising the most frequently occurring motifs.

(vi) Perform substructure searches for the component donor and acceptor fragments.

(vii) Combine results for donor/acceptor fragments and derive the number of possible occurrences of the motifs.

(viii) Calculate probabilities based on actual and possible frequencies of occurrence.

### Identification of potential donors and acceptors

The ability to locate intermolecular fragments of undefined topology (*i.e.* with variable intramolecular bond path lengths between H-bond donor and acceptor atoms) is not available in the released CSD System. Hence, the search program QUEST 3D was modified in order to provide this additional functionality.

Potential H-bond donors and acceptors were identified on the basis of atom type and the present analysis was restricted to N—H and O—H donors and N and O acceptors. Deuterium was treated as being equivalent to H for the purpose of motif recognition. The October 1996 release of the CSD (*ca.* 160 000 structure determinations) was used throughout.[21] The search was restricted to structures which were fully matched, error-free at the 0.02 Å level, had *R*-factor $\leqslant 10\%$, contained the elements C, H, D, N, O, S, P, B, Si and halogens only, and for which there was a perfect match between the chemical and crystallographic connectivity descriptions. Only structures for which H coordinates are stored were included in the analysis, avoiding any ambiguity as to the protonation state of N or O donor and acceptor atoms.

### Intermolecular contact search

Each donor and acceptor combination was considered as a potential H-bond and contacts X—H···A—Y were found using a modified version of the existing contact search routine, which is based on the algorithm described by Rollett.[22] To permit meaningful comparison of the H···A distances it was necessary to normalise the X—H distances to ideal values derived from neutron diffraction studies, placing H atoms along the existing X—H vector but at a standard distance of 1.009 and 0.983 Å for N—H and O—H bonds respectively.[23] Additionally, the condition that the X—H···A angle must be greater than 90° was imposed.

Whilst it is difficult to establish distance criteria for H-bonding, since the principal attractive (electrostatic) terms vary as $r^{-1}$ and $r^{-3}$, working distance limits are required for motif recognition based on crystallographic data. These were established using standard QUEST 3D non-bonded searches for intermolecular N—H···N, N—H···O, O—H···N and O—H···O contacts for which the angle X—H···A > 90°. Symmetry-equivalent fragments were rejected and contacts up to the sum of the appropriate van der Waals radii[24] were accepted. These searches yielded more than 10 000 fragments for N—H···O and O—H···O contacts and a limit of $R \leqslant 4\%$ was imposed for these, to permit examination of geometrical distributions in the CSD program VISTA. Visual inspection of the minima in the frequency distributions (Fig. 1) gave limits of 2.30 (N—H···N), 2.25 (N—H···O) and 2.20 Å (O—H···N, O—H···O). H-bond distance distributions are dependent on the chemical environment of the donor and acceptor, not the element type alone. Thus, ether and carbonyl O atoms are not equally good acceptors, and the histograms in Fig. 1 represent the superposition of several chemically-independent distributions. However, since the composite distributions are unimodal, the generalised distance limits above were deemed appropriate for this study.

The contact search was modified where one or more of the starting molecules possessed crystallographic symmetry. In such cases, additional symmetry-generated atoms are included in the CSD, so that the crystallographic connectivity description represents complete molecules. These complete molecules are related to their symmetry-generated neighbours in the extended crystal structure by more than one symmetry operation. When generating symmetry-related molecules in the intermolecular contact search, the redundant operations were not applied to avoid generating duplicate contacts.

For each X—H···A contact located, *i.e.* an H-bond from H in an original molecule to an acceptor A in a symmetry-

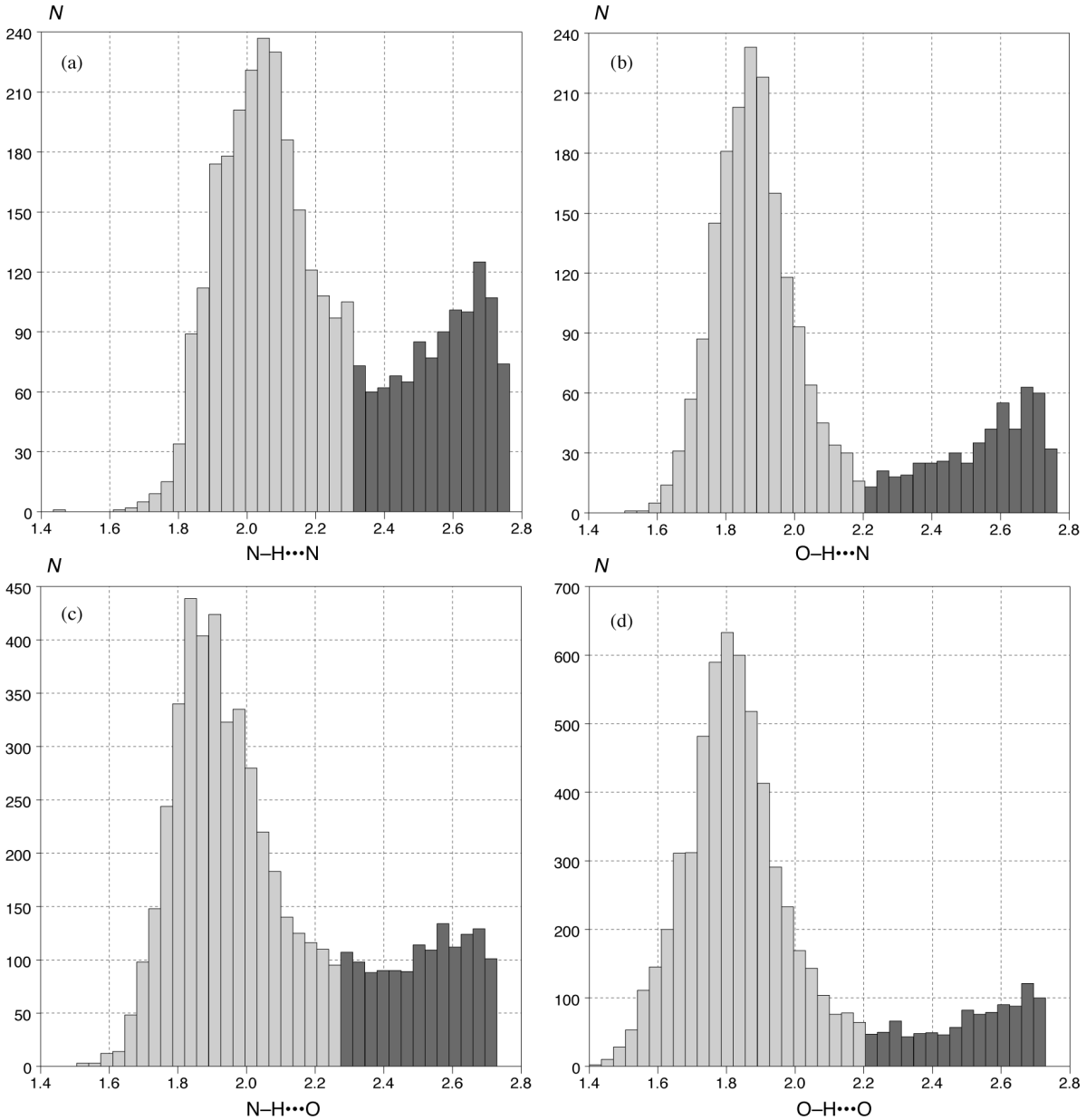


**Fig. 1** Histograms of (a) N–H···N, (b) O–H···N, (c) N–H···O and (d) O–H···O contact distances.

related or crystallographically-independent neighbouring molecule, the symmetry operation $\boldsymbol{S}$ applied to the acceptor molecule was stored. H-bonds from acceptors in the starting molecule(s) to donors in the neighbouring molecules were derived by considering the inverse symmetry operation $\boldsymbol{S}^{-1}$ relating the acceptor molecule back to its original position for each X—H···A contact found. Having established the H-bonds present, the intramolecular bond paths between all pairs of donors or acceptors in each starting molecule were derived. No attempt was made to account for duplicate paths in the recognition process; for small rings such paths are rare and, where they do occur, they are commonly chemically equivalent.

### Location of intermolecular rings

Each pair of starting molecules was considered in terms of its hydrogen bond contacts. If two molecules were joined by a single H-bond, the motif was taken as being a finite discrete motif and was neglected in the analysis. Where two molecules were joined by two or more H-bonds, each of the possible pairwise combinations of H-bonds constitutes a ring motif and was recorded separately. Bifurcated patterns were included, although they involve only one donor or acceptor in one of the molecules. No attempt was made to characterise explicitly motifs involving more than two H-bonds, by ring analysis or any other method, and only the individual components have been considered.

Where some molecules possess internal symmetry, symmetry-equivalent rings were retained. This is equivalent to treating the structure as if it were in a sub-group for which the internal symmetry operator of the higher symmetry molecule was absent. If a ring is formed between a molecule which has internal symmetry and a molecule which does not, the rings were also counted more than once, since the lower symmetry molecules would be crystallographically independent in the sub-group. This facilitates a more straightforward comparison with the number of possible motifs, $N_{\mathrm{poss}}$ [eqn. (1)], in any structure than if symmetry-equivalent motifs were rejected.

### Generation of graph set descriptors[16]

At this stage, the topology of a cyclic motif is now fully identified in terms of its ring size R($n$) and the numbers of acceptors (a) and donors (d) so that the appropriate descriptor, of the form $\mathrm{R}^{\mathrm{a}}_{\mathrm{d}}(n)$, can be assigned. For individual structures, the full

set of R-type descriptors (describing all cyclic motifs located in the extended structure) can be displayed. For our purposes, however, it was more efficient to generate the R-descriptors for the most frequently occurring motifs after completing the chemical classification and counting exercise described in the next section, *i.e.* they were generated from the chemical constitution keys used there.

The generation of graph-set descriptors for individual structures has recently been generalised to encompass the C, D and S-type descriptors[16] noted in the Introduction. This code has been incorporated in the CSD System structure visualiser, PLUTO, and displays graph sets up to the second level, subject to some restrictions for molecules which possess internal crystallographic symmetry. Full details of this implementation will be published shortly.[25]

### Classification and counting of motifs

Ring motifs were characterised on the basis of atom types (element and coordination number) and bond types (CSD bond type and cyclicity). Two amendments were made to the CSD bond type conventions to overcome ambiguities in connectivity coding: a 'guanidinium' bond type was introduced for cationic carbon bound to 3 three-coordinate nitrogen atoms (in the conventional CSD description two C—N bonds are described as single and one as double although they are equivalent chemically). A 'terminal oxygen' bond type for any mono-coordinate oxygen for which the bonds were described as single, double or delocalised was also introduced. This provides a consistent description of the bonding in carboxylate, phosphate and related anions. Similar ambiguities exist in the definitions of certain aromatic nitrogen moieties which may be coded with alternate single and double bonds; no attempt was made to take the alternative descriptions into account for such systems. For chemically symmetric motifs, a record was also made of whether or not the ring was formed about a crystallographic symmetry element. An upper limit of 20 atoms was placed on the ring size.

For each structure, every motif identified was compared with those already recorded in previous structures. A running total was kept of the number of occurrences of each discrete motif and the number of structures in which it was found. For every structure containing an H-bond motif, the motifs identified in that structure were logged in a file in motif-number order. Rather than comparing the explicit atom and bond properties of a new motif with those for all pre-existing motifs, atom and bond type integer keys were derived for each motif by packing the number of atoms and bonds of particular types. To increase the efficiency of searching on these two integer keys, they were arranged as nodes in a binary tree with no duplicate keys. A full list was made of the atom and bond properties for each motif corresponding to a particular key and these were compared in full once a key had been matched. By convention, a motif was defined as starting with an H atom and an H-bond. This leaves an ambiguity as to which H atom should be chosen if more than one donor is involved, making it necessary to test both possibilities when comparing motifs, reversing the atom and bond order where appropriate. In motifs involving a bifurcated hydrogen, the donor atom X does not comprise part of the ring; however, its identity is required if the chemical nature of the ring is to be defined and therefore it was included in the motif description.

To minimise any bias in the sample, motifs were recorded only once if more than one determination of a structure occurred in the CSD, only the first determination which exhibited a motif being retained. After the search, all discrete motifs were sorted both chemically (in terms of atom and bond keys) and according to the frequency with which they occurred. For each ring, the total numbers of motifs which were (a) associated with a symmetry element ($N_{sym}$: only possible if the motif is symmetric chemically) or (b) structurally asymmetric ($N_{asym}$), were also derived.

### Identification and counting of constituent covalent fragments

The total number of motifs $N_{poss}$ [eqn. (1)] which *could* be formed is dependent on the number of covalent fragments comprising the motif which are present in a structure. The first step in establishing $N_{poss}$ involves identifying the fragments comprising the motifs and performing standard QUEST 3D substructure searches for them. The motif recognition program generated an output file which described each motif in terms of its constituent atom and bond types, and which served as input to a stand-alone fragment recognition program. The motifs were sorted in terms of the number of structures in which they occurred and only those found in $>12$ structures were considered for the fragment identification process. Each motif was decomposed into its constituent covalent fragments, which were compared with those which had already been identified. The same atom and bond keys used to classify the motifs were used to define the constituent fragments and the fragments were described in full with the same atom and bond types. Where necessary, the fragments were reversed to achieve a match. This program produced a file containing the motif identifiers, motif statistics and the identifiers for the fragments comprising the motif.

For each fragment, a QUEST 3D connectivity search query was written automatically, comprising the required atoms (with total coordination number defined) and bonds. The QUEST 3D symmetry-check facility was set to retain symmetry-equivalent fragments, for consistency with the approach adopted in motif recognition. The same secondary search criteria and associated element tests used for motif recognition were adopted here, in order to search an identical subset of the CSD. Manual editing of the substructural queries was only necessary for the 'guanidinium' bonds, there being no equivalent in the CSD. Bonds to mono-coordinate oxygen were defined in substructural queries with a variable bond type setting (single, double or delocalised). Each query comprised a QUEST 3D instruction file, and the searches were run consecutively using a suitable script. For each query, a tables file was produced containing the (CSD) atom labels for each fragment and the structure identifier.

### Derivation of the number of possible motifs

The $N_{poss}$ values were derived, using a separate program, from the covalent fragments listed in these tables files. Duplicate determinations of the same crystal structure were omitted. Where some molecules in a structure contained symmetry-generated atoms, the structure was treated as if it were in a lower symmetry sub-group in which the operators generating the symmetry-atoms were absent, to be consistent with the method used to count the observed motifs.

The main difficulty in deriving $N_{poss}$ lies in determining which overlapping fragments could simultaneously form motifs. The lower limit is given by excluding all overlap and the upper limit by including all overlapping fragments. The problem is not merely topological, however, being complicated by the need to consider the three-dimensional geometric nature of the motif. Bifurcation, as well as other multi-centre H-bonds and H-bonds which are shared between two rings, must also be taken into account. A set of fragment combination rules (Fig. 2) was derived, which takes into account molecular symmetry, fragment overlap and the possibility of multi-centre H-bonding. These rules express which fragments cannot simultaneously form motifs, *i.e.* are mutually exclusive. For example, only one fragment is available from an amide $RC(=O)NH_2$ (Y same for both H's), reflecting the impossibility of both H's being *cis* to the carbonyl oxygen, and from a sulfonamide $RS(=O)_2NHR$ (X same for both A atoms), since

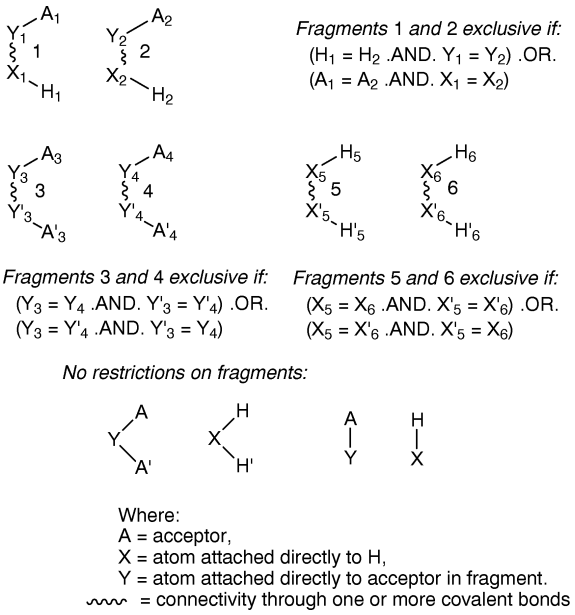


**Fig. 2** Combination rules for covalent fragments comprising motifs.

the H cannot be *cis* to both oxygen acceptors. Similarly, there are only three HNCNH fragments in a $[C(NH_2)_3]^+$ cation, with H atoms mutually *cis*, which may simultaneously form motifs.

The maximum number of chemically-equivalent fragments which could simultaneously form motifs was determined by considering all the fragment combinations permitted by these rules. No attempt was made to ensure that all the allowed fragments could be assembled into motifs; in some instances this is not possible because of the relative geometric disposition of the functional groups. Consider amide fragments having complementary functionality of two donors plus one acceptor and one donor plus two acceptors respectively [Fig. 3(a)]. In this case, it is not possible to assemble them with three H-bonds to form two *cis* amide eight-membered rings: the individual H—N—C=O fragments may not all form *cis* amide dimers unless the amide H makes two H-bonds [Fig. 3(b)], which also may not be geometrically possible. Similarly, it was assumed that the molecule was sufficiently flexible to adopt a conformation which permits motif formation, although this might not be the case if the conformation was constrained by ring closure (*e.g.* hydroxyl substituents of a pyranose sugar). As a result, $N_{poss}$ could be an over-estimate, albeit a more realistic value than would be obtained if all overlapping fragments were included.

For motifs composed of two fragments which were chemically identical, $N_{poss}$ is simply equal to the number of fragments which may simultaneously form motifs. A molecule containing two chemically identical fragments P and Q, may form either P–P′, Q–Q′ or P–Q′ and Q–P′ motifs (where a prime denotes a symmetry-related fragment in a neighbouring molecule). In both cases, two motifs would have been counted

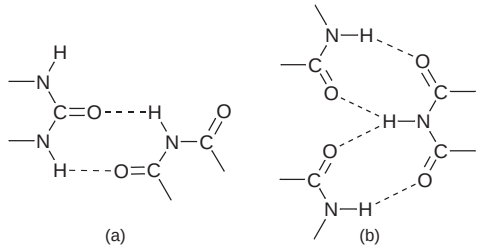


**Fig. 3** Complementary amide fragments (a) and bifurcated amide fragments (b).

in the motif search process, once for each fragment in the starting molecules, and the possible motifs are counted in the same manner. An overall count was made of the total number of possible motifs $N_{poss}$ in all structures and of the number of structures $S_{poss}$ that could possibly contain a motif.

Motifs which are not symmetric in terms of their chemical constitution were treated similarly, and the respective covalent fragments were read from two tables files concurrently. A structure was only considered further when it contained both chemically-distinct fragments. The number of covalent fragments of each chemical constitution which could form motifs simultaneously was derived. $N_{poss}$ is determined by the smaller of these two values: the number of fragments of the other chemical constitution which could form motifs has no effect on $N_{poss}$. Although the proportion by which one value exceeds the other could affect the probability of forming the motif, this was not taken into account in the derivation of $P_m$.

## Results

The 75 motifs having $S_{obs} > 12$ were ordered on their $P_m$-values and are depicted in Fig. 4; the motif numbers indicate the probability ranking of the motif. Relevant statistics for these motifs are given in Table 1, and graph sets[16] are also given for the motifs, in order to identify the more important topologies.

Motifs **1** and **2**, which have the highest probabilities of formation, are components of the triply hydrogen-bonded pattern shown in Fig. 5, which has some analogies with nucleotide recognition in DNA[26] and whose robustness in crystal engineering applications has been recognised by a number of authors.[27] Here, the individual rings are not independent and the motif might better be considered as the ensemble of the two ring motifs. The chemically-asymmetric $R_2^2(8)$ motif **2** occurs in more than 80% of possible structures and 90% of fragments. That $P_s$ is higher than $P_m$ suggests a degree of cooperativity, *i.e.* a second motif is more likely to form given the existence of one motif, although the difference may not be statistically significant. Whilst there are *ca.* 200 such fragments, these occur only in 29 structures. The $R_2^2(12)$ motif **1** represents the envelope of these rings; it occurs in every structure in which the respective fragments exist and $P_m$ is *ca.* 97%. These structures include many for which the planned use of this synthon in crystal engineering has been successful. It is proposed that the 1 : 1 complex of cyanuric acid and melamine contains infinite sheets connected by this motif[28] and the complex with 3(HCl) contains dimers linked by such units.[29] Its effectiveness can be attributed to the complementarity of the functional groups and the additional robustness engendered by three rather than two H-bonds.[5]

In terms of $P_m$, only 25 motifs have a probability greater than 30% (Table 1). The only motifs, in addition to **1** and **2**, with $P_m > 75\%$ are the carboxylic acid–oxime motif **3**, carboxylate–HNCNH motif **4** and carboxylic acid–HNCN ring **5**. None of these occur in more than 30 structures and all are chemically asymmetric. The $HNC^+NH$ fragment may be associated with carboxylate (**8**) and sulfate-type (**7**) functional groups and such motifs are formed in *ca.* 0.66 of the possible structures and for more than 50% of possible motifs. Motif **8** is commonly formed in biological systems between arginine and aspartic and glutamic acids.[26] The unusual motif **9**, incorporating two cyano-N···H-bonds, is formed with a probability of nearly 50% and is generally crystallographically symmetric (≈90%). Motifs **6** and **10**, components of crown ether host–guest motifs are two of the most probable. However, whilst **6** occurs in almost 85% of possible structures, $P_m$ is only 58%. This may be due to the manner in which the fragments are counted: although a $[NH_4]^+$ cation has six H–N–H combinations (allowing overlap) and the crown ether six (R)OCCOCCO(R) units, each ammonium ion may only

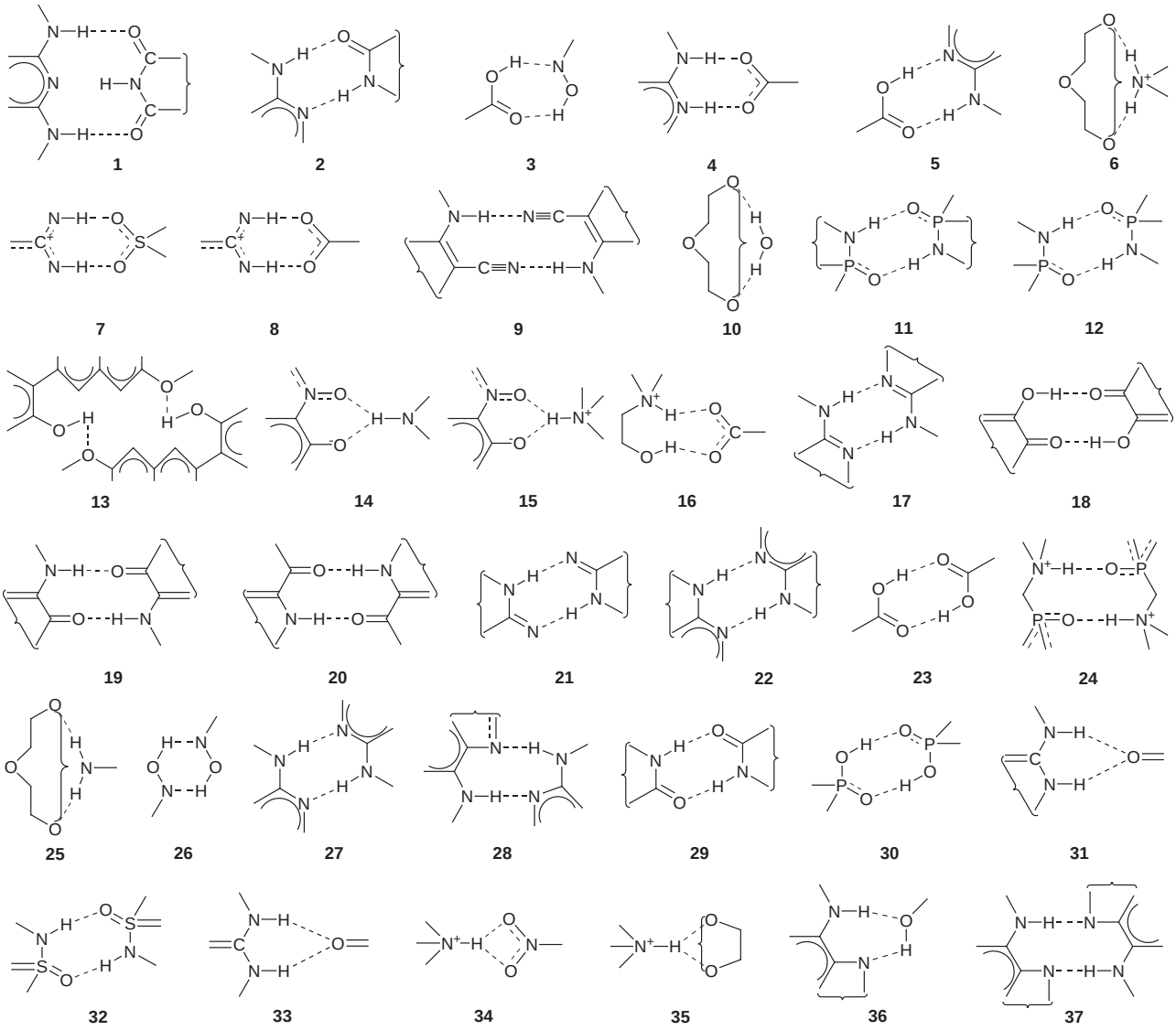


**Fig. 4** 75 H-bonded ring motifs occurring in >12 structures in the CSD.

form three H-bonds to a crown ether molecule (Fig. 6: the count will be given correctly as three if one of the substituents on the cationic N is not H). In these host–guest complexes, molecular recognition involves the three H atoms and the O atoms of the cyclic ether[30] and this pattern is best regarded as the H-bonding motif. Similar complexes with trivalent nitrogen (**25**) or water (**10**) are not as likely to occur as **6**, although they still exist in over half the number of possible structures.

The four most frequently occurring motifs include the cyclic $R_2^2(8)$ amide and carboxylic acid dimers. Perhaps surprisingly, the carboxylic acid dimer **23** only has a probability of *ca.* 0.33 of occurring, and the proportion of structures $P_s$ in which it occurs is similar. It has been suggested that aliphatic and aromatic carboxylic acids generally form the dimer motif,[9] although in some cases they form chains and a number of other patterns have been identified.[5] The relatively low overall probability may suggest that the motif is less prevalent than might have been thought, and this is due principally to the presence of competing functional groups, *e.g.* solvent water or carboxylate groups in partially deprotonated polycarboxylic acids. When restricted to those structures containing only C, O and H atoms, with no competing N, O acceptors or NH, OH donors in addition to carboxylic acid groups, the probabilities of formation are significantly higher (Table 2). For structures containing any number of molecules, each possessing only one COOH group, $P_m$ and $P_s$ are *ca.* 95% for the $R_2^2(8)$ motif, indicating a strong preference for the formation of intermolecular rings rather than chains. $P_m$ is somewhat lower for dicarboxylic acids (*ca.* 85%); this is due in part to the alternative possibility of forming an intramolecular S(7) ring in 1,2-dicarboxylic acids. Dicarboxylic acids have a lower probability of forming rings across a crystallographic symmetry element than monoacids (60 *vs.* 75%); in monoacids, the number of crystallographically independent molecules must be >1 if an asymmetric motif is to be formed, whereas such a motif may form between two COOH groups of the same diacid molecule provided it does not itself lie on a crystallographic symmetry element.

The lactam dimer **29**, which has the largest value of $N_{obs}$, occurs in a similar proportion of structures ($P_s$) to **23** although the $P_m$ figure is smaller (*ca.* 25%). This may suggest that some fragments are geometrically incompatible in certain structures, perhaps due to fragment overlap as discussed above (Fig. 3). The H donors are constrained to adopt the appropriate *cis* conformation if the ring size is sufficiently small, although the *trans* conformation is commonly adopted in larger rings (≥9 atoms) and this is incompatible with an $R_2^2(8)$ motif. In structures comprising only C, N, O and H atoms, with no competing H-bond donor/acceptor groups, the overall probability of 44% is less than half of $P_m$ for **23** (86%). The $P_m$ values for the $R_2^2(8)$ pattern are higher in mono- and di-amides where the HNC=O groups do not overlap (note that *ca.* 17% of the

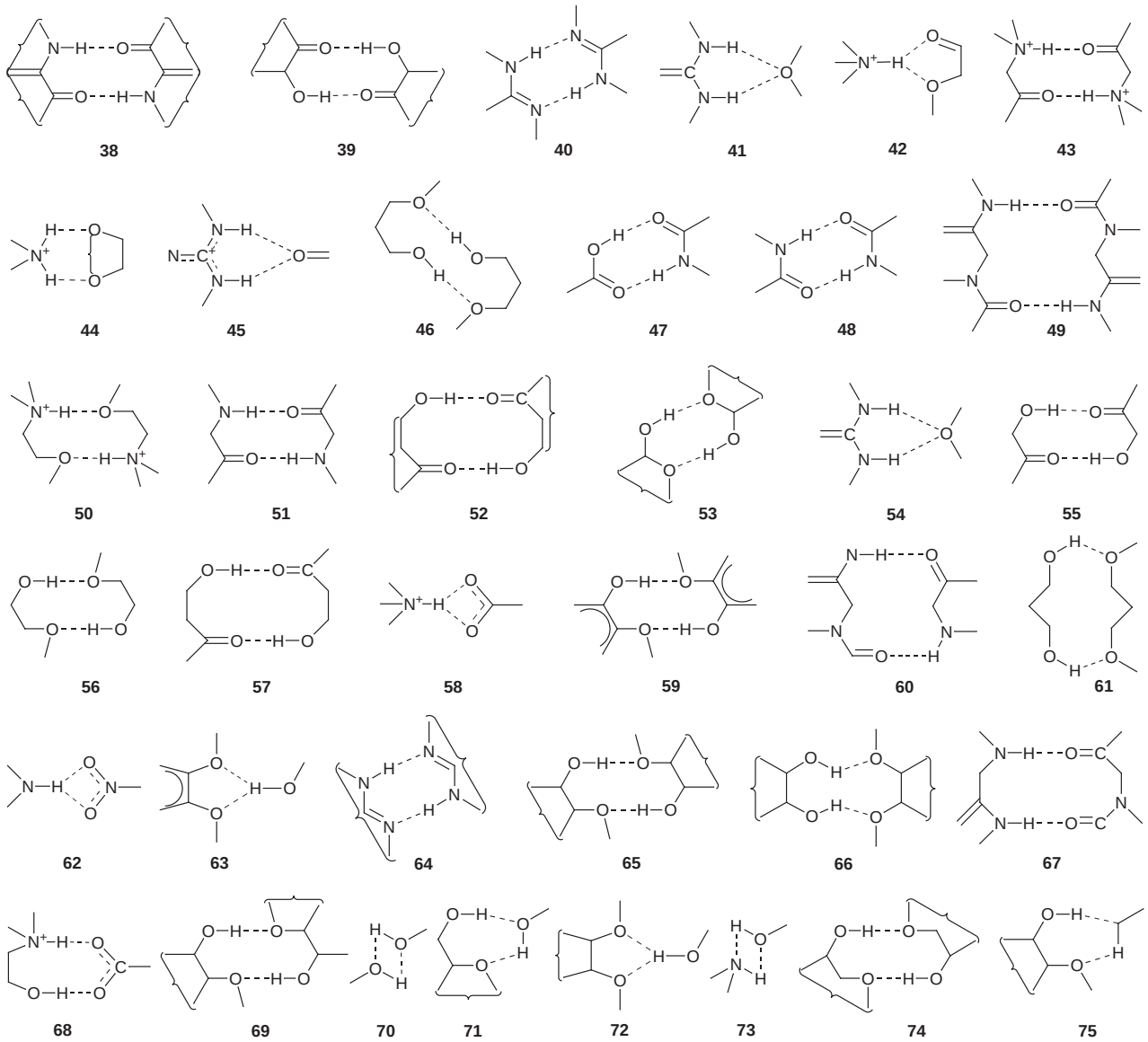

**Fig. 4** Continued

amide groups are *trans* in each case). For ureas HNC(=O)NH, the probabilities of formation are significantly higher based on the sample of eight structures; the motifs tend to be crystallographically symmetric and the molecules typically lie on a crystallographic symmetry element. In contrast, $P_m$ is only *ca.* 40% for O=CN(H)C=O molecules, since it is less favourable to adopt a bifurcated configuration about an NH donor than a C=O acceptor group. $P_s$ is rather higher (*ca.* 60%), since the diamide fragment may form one but not two rings in the absence of bifurcation at the donor. As a consequence, molecular symmetry is less likely to be reflected in the crystal due to the lower symmetry of the molecular environment.

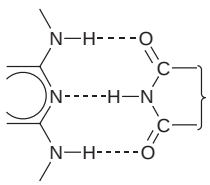

**Fig. 5** Motif formed with two fused rings of **2**.

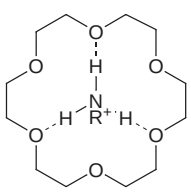

**Fig. 6** Ammonium/1,4,7,10,13,16-hexaoxacyclooctadecane (18-crown-6) motif.

The acyclic amide dimer **48** is much less likely to occur ($P_m < 10\%$) which may reflect an increased tendency to form chain motifs when the conformation of the amide is not constrained by cyclisation. $P_m$ is only 16% when the subset is restricted to molecules containing C, N, O and H with no competing donor groups. In secondary amides, the HNC=O groups are usually *trans* for steric and electronic reasons and form chains [the only *cis* example, which has a sterically undemanding H atom as the other C=O substituent, does form the $R_2^2(8)$ motif]. In contrast, ring motif $R_2^2(8)$ is much more prevalent in primary amides, since one of the donor H atoms is necessarily *cis*, and the probabilities $P_m$ and $P_s$ are in the range 80–95%, *i.e.* comparable to those for the carboxylic acid dimers **23** in the absence of strong competing H-bonding groups. The overall proportions of motifs **23**, **29** and **48** which form across a symmetry element are similar (each *ca.* 66%). In contrast, the cyclic sulfonamide dimer **32** (which has a <20%

**Table 1** Statistics for the 75 motifs with $S_{poss} \geqslant 12$

| Motif | Graph set | $P_m$(%) | $N_{obs}$ | $N_{poss}$ | Rank in $S_{obs}$ | $P_s$(%) | $S_{obs}$ | $S_{poss}$ | $P_{symm}$(%) |
|---|---|---|---|---|---|---|---|---|---|
| **1** | R2,2(12) | 97 | 93 | 96 | 16 | 100 | 25 | 25 | — |
| **2** | R2,2(8) | 91 | 199 | 218 | 7 | 83 | 29 | 35 | — |
| **3** | R2,2(7) | 90 | 36 | 40 | 48 | 92 | 12 | 13 | — |
| **4** | R2,2(8) | 82 | 36 | 44 | 49 | 81 | 17 | 21 | — |
| **5** | R2,2(8) | 76 | 62 | 82 | 29 | 71 | 20 | 28 | — |
| **6** | R2,2(10) | 58 | 206 | 354 | 5 | 84 | 42 | 50 | — |
| **7** | R2,2(8) | 54 | 158 | 290 | 9 | 63 | 26 | 41 | — |
| **8** | R2,2(8) | 51 | 79 | 154 | 21 | 67 | 26 | 39 | — |
| **9** | R2,2(12) | 47 | 21 | 45 | 70 | 47 | 18 | 38 | 91 |
| **10** | R2,2(10) | 45 | 86 | 192 | 18 | 62 | 34 | 55 | — |
| **11** | R2,2(8) | 43 | 20 | 46 | 74 | 59 | 17 | 29 | 80 |
| **12** | R2,2(8) | 41 | 38 | 92 | 47 | 44 | 30 | 68 | 68 |
| **13** | R2,2(20) | 39 | 15 | 38 | 75 | 76 | 13 | 17 | 100 |
| **14** | R2,1(6) | 39 | 45 | 114 | 40 | 36 | 14 | 39 | — |
| **15** | R2,1(6) | 39 | 47 | 120 | 38 | 38 | 17 | 45 | — |
| **16** | R2,2(9) | 37 | 44 | 118 | 43 | 42 | 20 | 48 | — |
| **17** | R2,2(8) | 37 | 204 | 556 | 6 | 37 | 159 | 435 | 75 |
| **18** | R2,2(10) | 36 | 58 | 159 | 32 | 38 | 47 | 123 | 62 |
| **19** | R2,2(10) | 36 | 20 | 55 | 73 | 34 | 16 | 47 | 60 |
| **20** | R2,2(10) | 35 | 39 | 111 | 45 | 35 | 30 | 86 | 64 |
| **21** | R2,2(8) | 35 | 29 | 83 | 57 | 40 | 25 | 63 | 79 |
| **22** | R2,2(8) | 33 | 27 | 81 | 63 | 37 | 21 | 57 | 63 |
| **23** | R2,2(8) | 33 | 847 | 2541 | 2 | 32 | 596 | 1873 | 65 |
| **24** | R2,2(10) | 32 | 23 | 71 | 66 | 47 | 17 | 36 | 57 |
| **25** | R2,2(10) | 32 | 99 | 306 | 13 | 55 | 31 | 56 | — |
| **26** | R2,2(6) | 27 | 93 | 341 | 17 | 28 | 72 | 256 | 76 |
| **27** | R2,2(8) | 26 | 172 | 660 | 8 | 32 | 126 | 390 | 64 |
| **28** | R2,2(9) | 24 | 50 | 206 | 35 | 26 | 23 | 89 | — |
| **29** | R2,2(8) | 24 | 876 | 3687 | 1 | 34 | 627 | 1796 | 62 |
| **30** | R2,2(8) | 21 | 84 | 404 | 19 | 29 | 59 | 205 | 64 |
| **31** | R1,2(6) | 18 | 54 | 296 | 33 | 19 | 21 | 113 | — |
| **32** | R2,2(8) | 17 | 61 | 350 | 30 | 21 | 53 | 253 | 87 |
| **33** | R1,2(6) | 17 | 153 | 904 | 10 | 18 | 63 | 344 | — |
| **34** | R2,1(4) | 16 | 74 | 450 | 25 | 14 | 21 | 150 | — |
| **35** | R2,1(5) | 16 | 44 | 268 | 41 | 16 | 15 | 92 | — |
| **36** | R2,2(7) | 16 | 30 | 192 | 56 | 18 | 14 | 80 | — |
| **37** | R2,2(10) | 16 | 22 | 141 | 68 | 15 | 18 | 119 | 73 |
| **38** | R2,2(10) | 13 | 26 | 196 | 65 | 15 | 20 | 132 | 62 |
| **39** | R2,2(10) | 12 | 59 | 488 | 31 | 14 | 54 | 390 | 83 |
| **40** | R2,2(8) | 11 | 28 | 258 | 59 | 13 | 17 | 136 | 43 |
| **41** | R2,2(6) | 10 | 31 | 297 | 54 | 9 | 21 | 223 | 74 |
| **42** | R2,1(5) | 10 | 32 | 312 | 51 | 13 | 16 | 122 | — |
| **43** | R2,2(10) | 10 | 78 | 781 | 23 | 13 | 55 | 431 | 54 |
| **44** | R2,2(7) | 10 | 48 | 486 | 36 | 22 | 17 | 79 | — |
| **45** | R1,2(6) | 10 | 78 | 790 | 22 | 19 | 24 | 126 | — |
| **46** | R2,2(12) | 10 | 74 | 773 | 27 | 19 | 37 | 193 | 11 |
| **47** | R2,2(8) | 10 | 67 | 702 | 28 | 9 | 26 | 282 | — |
| **48** | R2,2(8) | 8 | 361 | 4310 | 4 | 10 | 268 | 2614 | 66 |
| **49** | R2,2(14) | 8 | 48 | 637 | 37 | 9 | 29 | 330 | 33 |
| **50** | R2,2(10) | 8 | 21 | 279 | 69 | 7 | 14 | 207 | 71 |
| **51** | R2,2(10) | 7 | 101 | 1362 | 12 | 9 | 62 | 726 | 33 |
| **52** | R2,2(12) | 7 | 32 | 434 | 52 | 8 | 28 | 344 | 94 |
| **53** | R2,2(8) | 7 | 28 | 402 | 61 | 7 | 24 | 349 | 86 |
| **54** | R1,2(6) | 6 | 31 | 532 | 53 | 7 | 13 | 189 | — |
| **55** | R2,2(10) | 5.4 | 28 | 519 | 60 | 6 | 21 | 355 | 64 |
| **56** | R2,2(10) | 5.4 | 80 | 1490 | 20 | 10 | 48 | 483 | 30 |
| **57** | R2,2(12) | 5.1 | 21 | 412 | 71 | 7 | 20 | 282 | 100 |
| **58** | R2,1(4) | 5.1 | 96 | 1892 | 15 | 6 | 42 | 708 | — |
| **59** | R2,2(10) | 4.6 | 26 | 570 | 64 | 6 | 17 | 275 | 54 |
| **60** | R2,2(12) | 4.4 | 54 | 1234 | 34 | 7 | 23 | 314 | — |
| **61** | R2,2(12) | 4.3 | 32 | 750 | 50 | 8 | 13 | 156 | — |
| **62** | R2,1(4) | 3.8 | 74 | 1936 | 26 | 4.3 | 30 | 691 | — |
| **63** | R2,1(5) | 3.8 | 44 | 1160 | 42 | 4.9 | 20 | 407 | — |
| **64** | R2,2(8) | 3.7 | 21 | 567 | 72 | 4.5 | 20 | 442 | 100 |
| **65** | R2,2(10) | 3.7 | 135 | 3702 | 11 | 7 | 79 | 1121 | 48 |
| **66** | R2,2(10) | 3.2 | 97 | 3002 | 14 | 5.5 | 47 | 861 | — |
| **67** | R2,2(12) | 3.1 | 38 | 1232 | 46 | 5.7 | 18 | 316 | — |
| **68** | R2,2(9) | 2.9 | 40 | 1378 | 44 | 3.9 | 19 | 488 | — |
| **69** | R2,2(10) | 2.2 | 28 | 1256 | 58 | 3.1 | 14 | 458 | — |
| **70** | R2,2(4) | 1.9 | 408 | 21824 | 3 | 1.9 | 203 | 10694 | 14 |
| **71** | R2,2(7) | 1.1 | 27 | 2466 | 62 | 1.5 | 14 | 907 | — |
| **72** | R2,1(5) | 1.0 | 46 | 4644 | 39 | 1.9 | 22 | 1173 | — |
| **73** | R2,2(4) | 1.0 | 75 | 7682 | 24 | 1.2 | 32 | 2768 | — |
| **74** | R2,2(10) | 0.8 | 22 | 2671 | 67 | 1.3 | 18 | 1369 | 64 |
| **75** | R2,2(7) | 0.4 | 30 | 7048 | 55 | 1.2 | 13 | 1121 | — |

**Table 2** Statistics for motifs **23**, **29** and **48** in structures with no competing donors or acceptors

| Motif | | $P_m$(%) | $N_{obs}$ | $N_{poss}$ | $P_s$(%) | $S_{obs}$ | $S_{poss}$ | $P_{symm}$(%) |
|---|---|---|---|---|---|---|---|---|
| **23** | All | 86 | 311 | 361 | 91 | 181 | 198 | 64 |
| | Mono | 96 | 128 | 134 | 95 | 105 | 111 | 75 |
| | Di | 85 | 139 | 164 | 88 | 63 | 72 | 60 |
| **29** | All | 44 | 136 | 308 | 64 | 72 | 114 | 74 |
| | Mono | 52 | 17 | 33 | 52 | 15 | 29 | 88 |
| | Di, no overlap | 57 | 33 | 58 | 59 | 17 | 29 | 70 |
| | Di, shared NH | 86 | 19 | 22 | 100 | 8 | 8 | 90 |
| | Di, shared C=O | 40 | 15 | 38 | 62 | 8 | 13 | 73 |
| **48** | All, $1^y$ and $2^y$ | 16 | 50 | 318 | 22 | 35 | 158 | 84 |
| | Mono, $1^y$ | 83 | 19 | 23 | 90 | 17 | 19 | 79 |
| | Di, $1^y$ | 95 | 19 | 20 | 90 | 9 | 10 | 100 |
| | Mono, $2^y$ | 2.1 | 1 | 47 | 2.4 | 1 | 41 | — |
| | Di, $2^y$ | 0 | 0 | 52 | 0 | 0 | 46 | — |

probability of occurring) forms across a symmetry element in *ca.* 87% of cases.

Interestingly, the mixed amide/carboxylic acid dimer **47** is only as likely to occur ($P_m$ *ca.* 9%) as the chemically-symmetric acyclic amide motif **48**, again reflecting the tendency of acyclic amides to adopt the *trans* conformation. The $R_2^2(8)$ dimers **17** and **27** are similar chemically, with N donors and acceptors, differing only in whether the cyclic C–N bond is double or aromatic. For some structures, the precise identification of the motif as **17** or **27** will depend on the particular manner in which the structure has been coded in the CSD. **17** has a slightly higher probability of occurring than the $R_2^2(8)$ dimers **23**, **29** and **48** whereas **27** has $P_m$ closer to 0.33. **17** has a greater proportion of occurrences associated with a symmetry element (*ca.* 75% compared to 62–66% for **23**, **27**, **29** and **48**). As with motif **29**, **27** is found in a lower proportion of fragments ($P_m$) than structures ($P_s$) which may be due in part to a lack of fragment complementarity.

Interestingly, the cyclic (**11**) and acyclic (**12**) phosphoramide $R_2^2(8)$ dimers are more likely to form than either their amide or carboxylic acid analogues ($P_m > 40\%$); **11** is particularly likely to be associated with a crystallographic symmetry element ($P_{symm} = 80\%$). The most probable motifs involving bifurcated H atoms are $R_1^2(6)$ rings in which an ammonium (**15**), or amine (**14**), H is chelated by a nitro and aryloxy oxygen atom; these would appear to be significantly more probable than the more common motifs **34** and **62** in which the H is chelated by both nitro oxygen atoms.

The motif containing a bifurcated H which occurs in the largest number of structures comprises an ammonium cation and a carboxylate anion **58** (*e.g.* in amino acid zwitterions); this motif has only a *ca.* 5% probability of being formed. The analogous $R_1^2(4)$ motif with carboxylate replaced by nitro **34** is more than twice as likely to occur. However, if the ammonium is also replaced by tri-coordinate nitrogen the probability of forming **62** (*e.g.* in nitroanilines[2]) is only *ca.* 4%. However, it is possible that the distance limits used were too short for the longer contact to be found for **34** and **62** in cases where the chelation was more asymmetric. Of those in which two donor atoms chelate an acceptor, **33** occurs most frequently although $P_m$ is less than 20%. This $R_2^1(6)$ motif occurs typically in the 'head-to-tail' packing of diarylurea molecules.[31] The motif is formed with a similar probability if one of the C–N bonds is cyclic (**31**) although it is much less likely to occur if the oxygen atom is two-coordinate (**54**). Motif **45** (analogous to **33** but with a cationic three-coordinate C atom) occurs in the same proportion of structures as **33** although it only forms for on average 50% of the possible motifs in any structure in which it occurs, indicating that it is not possible to use all fragments simultaneously.

In contrast to the cyclic and acyclic amides, the chemically-symmetric HNCC=O dimer **51** is crystallographically-symmetric in only 0.33% of the motifs although it occurs with a similar probability to **48** ($P_m < 8\%$). If the amino group is replaced by cationic nitrogen, the motif (**43**) is somewhat more probable ($P_m \approx 10\%$) and is almost equally likely to be symmetric crystallographically as not. Alternatively, if the C–N bonds are cyclic (**20**) the H donor is constrained to a *cis* conformation and the ring motif occurs in a comparable proportion of structures ($P_s$) to that of the cyclic amide **29** and is more likely to be symmetric crystallographically.

The $HOC(R)_2C(R)_2OH$ unit may form either the symmetric motif **65** or asymmetric motif **66** (where the C–C bond is cyclic, *e.g.* in pyranose and furanose sugars) due to the ability of the O-hydroxyl atoms to act as both donor and acceptor; these motifs may also be adopted if some of the atoms are O-ether rather than O-hydroxyl. Interestingly, the probability of the chemically-symmetric motif **65** being crystallographically symmetric is close to 50%. The statistics for these motifs are based on the assumption that O-hydroxyl may act as both donor and acceptor [*i.e.* there are two fragments associated with a $HOC(R)_2C(R)_2OH$ unit]; whilst this does occur, it is relatively uncommon so that $P_m$ is about half the value of $P_s$ (which is less than 10% for both **65** and **66**). The analogue of **65** in which the C–C bonds are intramolecularly acyclic (**56**) is slightly more probable, perhaps because the conformation is not restricted by the constraints of ring closure.

Motif **70** represents a $R_2^2(4)$ tandem RO–H/H–OR hydrogen bond configuration; the large number of occurrences are due to the large number of RO–H fragments rather than any strong tendency to form the motif ($P_m < 2\%$). The majority of these motifs (>85%) are not associated with a crystallographic symmetry element. Calculations on the water dimer have shown that the tandem configuration is energetically disfavoured with respect to the $C_s$ terminal configuration by *ca.* 4 kJ $mol^{-1}$.[32] The $R_2N$–H/H–OR dimer **73** has an even lower probability of occurring (*ca.* 1%).

## Conclusions

A new methodology has been used to identify the most common rings involving two H-bonds which are formed between pairs of organic molecules in the CSD, and to calculate their relative probabilities of occurrence. Whilst the derivation of $N_{poss}$ might be refined further, comparison with the structural probability ($P_s$) indicates that the $P_m$ values obtained are reasonable (at least for the 75 motifs with $S_{obs} > 12$), and discrepancies may be rationalised on the basis of the geometrical characteristics of the constituent fragments.

Some motifs are less likely to occur in the presence of competing intermolecular interactions than might have been thought [*e.g.* carboxylic acid and *cis*-amide $R_2^2(8)$ dimers], a result that has clear consequences for crystal design. However, some of the motifs which have been applied specifically in crystal engineering (*e.g.* components of motifs comprising

three H-bonds), and host–guest complexes of crown ethers, have a high $P_m$, as expected. The survey has not revealed many other motifs with $P_m > 50\%$, although those with $P_m > 30\%$ could be of practical use in the absence of strong competing interactions.

These results have important implications, not only for crystal engineering, but also for molecular modelling and crystal structure prediction, and investigations are continuing on the effects of competing interactions on motif formation. The methodology can be applied to systems involving weaker H-bonds and other non-covalent interactions, including those in organometallic and inorganic systems. Work in these areas is in progress.

## Acknowledgements

The authors wish to thank Dr. R. Scott Rowland (CCDC) for the non-bonded search routine, Dr. Jason Cole (CCDC) for the script to run the fragment searches and Profs. Gautam Desiraju (Hyderabad), Joel Bernstein (Beer Sheva) and Ray Davis (Austin) for valuable discussions.

## References

1 G. A. Jeffrey, *An Introduction to Hydrogen Bonding*, Oxford University Press, New York, NY, 1997.
2 T. W. Panunto, Z. Urbánczyk-Lipkowska, R. B. Johnson and M. C. Etter, *J. Am. Chem. Soc.*, 1987, **109**, 7786; M. C. Etter and G. M. Frankenbach, *Materials*, 1989, **1**, 10; M. C. Etter, *J. Phys. Chem.*, 1991, **95**, 4601.
3 W. Jones, V. R. Pedireddi, A. P. Chorlton and R. Docherty, *Chem. Commun.*, 1996, 997.
4 F. Garcia-Tellado, S. J. Geib, S. Goswani and A. D. Hamilton, *J. Am. Chem. Soc.*, 1991, **113**, 9265; J. Bernstein, M. C. Etter and L. Leiserowitz, in *Structure Correlation*, ed. H.-B. Bürgi and J. D. Dunitz, VCH, Weinheim, 1994, vol. 2.
5 G. R. Desiraju, *Crystal Engineering—The Design of Organic Solids*, Elsevier, Amsterdam, 1989; *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 2311; *Chem. Commun.*, 1997, 1475.
6 C. B. Aakeröy and K. R. Seddon, *Chem. Soc. Rev.*, 1993, **22**, 397.
7 V. R. Thalladi, B. S. Goud, V. J. Hoy, F. H. Allen, J. A. K. Howard and G. R. Desiraju, *Chem. Commun.*, 1996, 401.
8 G. A. Jeffrey and S. Takagi, *Acc. Chem. Res.*, 1978, **11**, 264.
9 L. Leiserowitz and G. M. J. Schmidt, *J. Chem. Soc. A*, 1969, 2372.
10 L. Leiserowitz, *Acta Crystallogra., Sect. B*, 1976, **32**, 775; L. Leiserowitz and M. Tuval, *Acta Crystallog., Sect. B*, 1978, **34**, 1230.
11 A. F. Wells, *Structural Inorganic Chemistry*, Oxford University Press, Oxford, 1962.
12 G. Gilli, F. Belluci, V. Ferretti and V. Bertolasi, *J. Am. Chem. Soc.*, 1989, **111**, 1023.
13 W. F. Bailinger, P. v. R. Schleyer, T. S. S. R. Murty and L. Robinson, *Tetrahedron*, 1964, **20**, 1635.
14 G. A. Jeffrey, M. E. Gress and S. Takagi, *J. Am. Chem. Soc.*, 1977, **99**, 609.
15 L. N. Kuleshova and P. M. Zorkii, *Acta Crystallogr., Sect. B*, 1980, **36**, 2113; P. M. Zorkii and L. N. Kuleshova, *Zh. Strukt. Khim.*, 1980, **22**, 153.
16 M. C. Etter, J. C. MacDonald and J. Bernstein, *Acta Crystallogr., Sect. B*, 1990, **46**, 256; M. C. Etter, *Acc. Chem. Res.*, 1990, **23**, 120; J. Bernstein, R. E. Davis, L. Shimoni and N.-L. Chang, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 1555.
17 J. Bernstein, *Acta Crystallogr., Sect. B*, 1991, **47**, 1004.
18 J. Bernstein, M. C. Etter and J. M. MacDonald, *J. Chem. Soc., Perkin Trans. 2*, 1990, 695.
19 J. Bernstein and R. E. Davis, *Graph Set Analysis of Hydrogen Bond Motifs*, in *Implications of Molecular and Materials Structure for New Technologies*, ed. J. A. K. Howard and F. H. Allen, Kluwer Academic Publishers, Dordrecht, to be published.
20 F. H. Allen and O. Kennard, *Chem. Des. Automat. News*, 1993, **8**, 1; 31.
21 F. H. Allen, P. R. Raithby, G. P. Shields and R. Taylor, *Chem. Commun.*, 1998, 1043.
22 J. S. Rollett, in *Computing Methods in Crystallography*, ed. J. S. Rollett, Pergamon, Oxford, 1965.
23 F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen and R. Taylor, *J. Chem. Soc., Perkin Trans. 2*, 1987, S1.
24 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441.
25 F. H. Allen, W. D. S. Motherwell and G. P. Shields, *Acta Crystallogr., Sect. B*, submitted.
26 W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984, ch. 6; G. A. Jeffrey and W. Saenger, *Hydrogen Bonding in Biological Structures*, Springer-Verlag, Berlin, 1991.
27 J.-M. Lehn, M. Mascal, A. DeCian and J. Fischer, *J. Chem. Soc., Chem. Commun.*, 1990, 479; J. A. Zerkowski, C. T. Seto, D. A. Wierda and G. M. Whitesides, *J. Am. Chem. Soc.*, 1990, **112**, 9025; N. Kimizuka, T. Kawasaki and T. Kunitake, *J. Am. Chem. Soc.*, 1993, **115**, 4367; R. Ahuja, P.-L. Caruso, D. Möbius, W. Paulus, H. Ringsdorf and G. Wildberg, *Angew. Chem., Int. Ed. Engl.*, 1993, **32**, 1033; D. A. Bell and E. V. Anslyn, *J. Org. Chem.*, 1994, **59**, 512; G. M. Whitesides, E. E. Simanek, J. P. Mathais, C. T. Seto, D. N. Chin, M. Mammen and D. M. Gordon, *Acc. Chem. Res.*, 1995, **28**, 37.
28 C. T. Seto and G. M. Whitesides, *J. Am. Chem. Soc.*, 1993, **115**, 905.
29 Y. Wang, B. Wei and Q. Wang, *J. Cryst. Spectrosc. Res.*, 1990, **20**, 79.
30 M. S. Fonar, Yu. A. Simonov, A. A. Dvorkin, T. I. Malinovskii, E. V. Ganin, S. Kotlyar and V. F. Makarov, *J. Incl. Phenom.*, 1992, **12**, 3291; F. Seel. N. Klein, B. Krebs, M. Dartmann and G. Henkel, *Z. Anorg. Allg. Chem.*, 1985, **524**, 95.
31 M. C. Etter and T. W. Panunto, *J. Am. Chem. Soc.*, 1988, **110**, 5896.
32 M. J. Frisch, J. A. Pople and J. E. Del Bene, *J. Phys. Chem.*, 1985, **89**, 3664.

*Paper 8/07212D*