Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Hydrogen-bond motifs in the crystals of hydrophobic amino acids

László Fábián,[a,b]* James A. Chisholm,[b] Peter T. A. Galek,[a,b] W. D. Samuel Motherwell[b] and Neil Feeder[c]

[a]Pfizer Institute for Pharmaceutical Materials Science, Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, England, [b]Cam-Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England, and [c]Pfizer Global R&D, Ramsgate Road, Sandwich, Kent CT13 9NJ, England

Correspondence e-mail: fabian@ccdc.cam.ac.uk

Received 14 April 2008
Accepted 24 June 2008

A computer program has been developed to survey a set of crystal structures for hydrogen-bond motifs. Possible ring and chain motifs are generated automatically from a user-defined list of interacting molecular fragments and intermolecular interactions. The new program was used to analyse the hydrogen-bond networks in the crystals of 52 zwitterionic α-amino acids. All the possible chain motifs (repeating 1–4 molecules) are frequent, while the frequency of ring motifs (2–6 molecules) ranges from 0 to 85% of the structures. The list of motifs displayed by each structure reveals structural similarities and it can be used to compare polymorphs. The motifs formed in cocrystals of α-amino acids and in crystals of β- and γ-amino acids are similar to those of α-amino acids.

1. Introduction

Molecular crystals are held together by intermolecular interactions. Unit-cell translations and crystal symmetry combine individual interactions to form extended networks. Such networks often have common subsets in crystals of related compounds or in polymorphs (Fábián & Kálmán, 2004; Gelbrich & Hursthouse, 2005), *i.e.* they often display common interaction motifs. Recurrent motifs may indicate favourable interactions or the kinetic accessibility of a particular arrangement. The identification of frequent motifs is thus useful in crystal engineering (*e.g.* Vishweshwar *et al.*, 2003; Fábián *et al.*, 2005), *ab initio* crystal structure prediction (Nowell & Price, 2005; Cruz Cabeza *et al.*, 2006) and in understanding polymorphism (Blagden & Davey, 2003).

Hydrogen bonds have a distinguished role in governing crystallization because of their strength and directionality (Jeffrey, 1997). The systematic description of hydrogen-bond motifs has been enabled by the graph-set approach (Etter, 1990; Bernstein *et al.*, 1995; Grell *et al.*, 1999). The graph-set notation of a motif, $X_d^a(n)$, gives the type of the motif ($X = D$, C, R or S for discrete motif, infinite chain, ring or intramolecular ring motifs, respectively), the number of donor (d) and acceptor (a) atoms, and the total number of atoms comprising the motif (n). These topological descriptors have proven useful in describing the hydrogen-bonding preferences of several classes of compounds, *e.g.* polyamides (Motherwell *et al.*, 2000*a*), sulfonamides (Adsmond & Grant, 2001) and organic sulfonates (Haynes *et al.*, 2004).

In this work we report a systematic analysis of hydrogen-bond motifs found in crystal structures of small-molecule amino acids found in the Cambridge Structural Database (CSD; Allen, 2002). In these structures charge-assisted ammonium–carboxylate hydrogen bonds form a diverse range of complex network topologies. Finding common motifs can highlight the relationship between structures and thus can

contribute to a better understanding of their formation. To begin, 52 α-amino acid structures were surveyed that contain no other hydrogen-bonding groups than the ammonium and carboxylate groups. The results found for α-amino acid structures were then compared with those for β- and γ-amino acid structures.

2. Search method

Searches were performed using the algorithm *3DSearch* (Chisholm & Motherwell, 2004). This algorithm enables the efficient searching of crystal structure data and can readily identify extended motif search queries defined by chemical and geometric constraints. The careful selection of constraints used to define motifs is important as these determine the motif frequencies obtained.

One approach is to manually define motif search queries, for example by using a sketcher program to sketch the possible topological combinations of groups. However, this can be a time-consuming process which is prone to error and one can never be sure that a complete set of motif possibilities has been defined. To overcome this difficulty we have developed an algorithm, called 'motif factory', to automatically generate motif possibilities from an input list of one or more functional groups, and a set of close contacts specified between atoms of these groups. The algorithm assembles the component groups and contacts in all possible combinations or sequences to generate dimer motifs and ring and chain motifs of specified sizes. The size of the motif refers to the number of molecules spanned by the motif or, alternatively, the number of intermolecular contacts. For example, one can define groups A and B, define contacts A–A and A–B and generate all possible ring motifs that span four molecules (Fig. 1). Each generated motif is then distinguishable by having a distinct sequence of contact atoms either along the chain or around the ring. For example, the rings in Fig. 1 can be described by the strings (···A···A···A···A···), (···A···A···A···B···) and (···A···B···A···B···), respectively.[1]

It is appropriate to discuss some of the differences between the current search method and performing a graph-set analysis by an automated routine, such as that available in the *RPluto* program (Motherwell *et al.*, 1999, 2000*b*). The essential difference is that the graph-set method does not define motifs in terms of specific chemical fragments. The intention is rather to describe the topology of a hydrogen-bonded array. For example, the graph-set $R_2^2(8)$ can refer to both a carboxylic acid dimer motif and to a carboxamide dimer motif. Subsequent analysis of graph sets can, of course, be carried out to further classify motifs by chemical fragment, *e.g.* to identify isographic relationships between functional groups (Etter, 1990). However, no automated routine currently exists for this purpose and such a two-step process can be less efficient when analysing a large number of structures. The current method is useful when investigating the propensity of certain chemical fragments to form certain motifs rather than the propensity of the formation of certain graph sets.

A further difference is that, with the graph-set method, the symmetry relationship of the hydrogen bonds is included in the motif definition through the distinction of primary, secondary and higher-level graph sets. The current method ignores symmetry (except when determining whether chains are infinite). This overcomes the problem in *RPluto* which is limited to finding graph sets of secondary level, *i.e.* motifs composed of up to two symmetry-independent hydrogen bonds. By ignoring symmetry we can also identify common motifs in structures that have different space-group symmetries. One difficulty with the current approach and the use of the motif factory is that the number of motifs can quickly get out of hand as the number of groups and number of contacts and the size of motifs is increased. Also the nomenclature, although more descriptive, is not as concise as the graph-set notation. Given the obvious utility of the graph-set notation, for each motif found we also present the corresponding graph-set notation, but note that the frequencies presented are not graph-set frequencies.

The motif factory and the motif search functionalities have been implemented as a new module of the *Mercury* crystal structure visualiser (Macrae *et al.*, 2008). An intuitive graphical user interface is available to automatically generate motifs, to perform motif surveys and to analyse the results (Macrae *et al.*, 2008).

3. α-Amino-acid survey

The motif search algorithm was first tested on crystal structures of simple amino acids that contain no hydrogen-bonding groups other than the zwitterionic pair of ammonium and carboxylate groups. The Cambridge Structural Database (CSD, V5.29, January 2008 updates; Allen, 2002) was searched for structures of such α-amino acids. Only one-component crystals were considered, *i.e.* salts, solvates and cocrystals were excluded from the initial survey. After removing redeterminations, 52 matching crystal structures remained (Table 1).

The carboxylate group was defined by allowing both carbon–oxygen bonds to match any bond type. The total number of atoms bonded to each O atom was constrained to one. The ammonium group was defined by requiring the number of H atoms on the N atom to be three and the assigned atomic charge to be +1.

The only contact considered was the ammonium–carboxylate hydrogen bond, defined with an N···O distance cutoff of 3 Å between N(ammonium) and either O atom of a

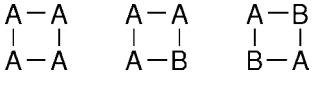

Figure 1
Possible four-membered rings formed by groups A and B with contacts A–A and A–B (*e.g.* 'A' could be a water molecule, and 'B' a carbonyl group).

[1] This notation of the ring motifs is not unique. Any circular permutation of the same molecular fragments represents the same ring.

Table 1
α-Amino acid structures in the CSD without hydrogen-bonding groups other than the ammonium and carboxylate groups.

Unit-cell dimensions are given in the supplementary material.

REFCODE	Name	Space group	Z
ACXTPY	4-Amino-thiapyran-4-carboxylic acid	$P2_1/c$	4
ACYHXA01	1-Amino-cyclohexane-1-carboxylic acid	$P2_1/a$	4
AHEJEC01	D-Valine	$P2_1$	4
ALUCAL04	D-Alanine	$P2_12_12_1$	4
AMMPRA01	α-Ammonium α-methylpropionoate	$C2/c$	8
BOQCUF	DL-Cysteine	$P2_1/a$	4
CEDFAS	*rac*-DL-Penicillamine	$P2_1/c$	4
CERGEL	DL-Selenomethionine	$P2_1/a$	4
CHEDGL10	*rac*-1,4-Cyclohexadiene-1-glycine	$P2_1/c$	4
DAILEU	D-Alloisoleucine	$P2_1$	4
DEYFOD	L-Selenomethionine	$P2_1$	4
DLABUT03	DL-α-Amino-*n*-butyric acid (form D)	$I2/a$	8
DLABUT05	DL-α-Amino-*n*-butyric acid (form A)	$P2_1/a$	4
DLABUT11	DL-α-Amino-*n*-butyric acid (form B)	$P4_2/n$	8
DLALNI01	DL-Alanine	$Pna2_1$	4
DLILEU02	DL-Isoleucine	$P\bar{1}$	2
DLLEUC	DL-Leucine	$P\bar{1}$	2
DLMETA02	DL-Methionine (α form)	$P2_1/a$	4
DLMETA05	DL-Methionine (β form)	$I2/a$	8
DLNLUA01	DL-Norleucine (α form)	$P2_1/a$	4
DLNLUA02	DL-Norleucine (β form)	$C2/c$	8
EXAXEG	4-Fluorophenylalanine	$P2_1$	4
EXAXOQ	3,4,5-Trifluorophenylalanine	$P2_1$	4
FEGMAG	β-Chloro-L-alanine	$P2_12_12_1$	4
FIVGEW	DL-2-Amino-2-phenylethanoic acid	$P2_1/c$	4
GLYCIN16	Glycine (γ form)	$P3_2$	3
GLYCIN19	Glycine (α form)	$P2_1/n$	4
GLYCIN27	Glycine (β form)	$P2_1$	2
GLYCIN36	Glycine (δ form)	Pn	2
GLYCIN68	Glycine (ε form)	Pn	2
ICAMOO	*S*-Benzyl-L-cysteine	$P2_1$	4
JOXBED	*S*-(2,2,2-Trifluoroethyl)-L-cysteine	$P2_1$	4
KEPJOE	(±)-2-*exo*-Amino-6-*endo*-(methylthio)bicyclo[2.2.1]heptane-2-*endo*-carboxylic acid	$P2_1/c$	4
LALNIN12	L-Alanine	$P2_12_12_1$	4
LALXAX	*rac*-2-Amino-2-(2-fluorophenyl)-acetic acid	$P2_1/c$	4
LARSUT	*rac*-2-Amino-4-fluorododec-4-enecarboxylic acid	$P2_1/c$	4
LCYSTN04	L-Cysteine [form (II)]	$P2_1$	4
LCYSTN21	L-Cysteine [form (I)]	$P2_12_12_1$	4
LCYSTN24	L-Cysteine [form (III)]	$P2_12_12_1$	4
LCYSTN26	L-Cysteine [form (IV)]	$P2_1$	4
LEUCIN01	L-Leucine	$P2_1$	4
LEUCIN03	L-Leucine (high-pressure form)	$C2$	4
LISLEU02	L-Isoleucine	$P2_1$	4
LMETON02	L-Methionine	$P2_1$	4
LNLEUC10	L-Norleucine	$C2$	4
LVALIN01	L-Valine	$P2_1$	4
SIMPEJ	(*R*)-Phenylalanine	$C2$	8
SIMRAH	*(R,S)*-Hexafluorovaline	$P2_1/c$	4
VALIDL	DL-Valine (monoclinic)	$P2_1/c$	4
VALIDL02	DL-Valine (triclinic)	$P\bar{1}$	2
XADTUR	DL-2-Amino-4-pentenoic acid	$Pca2_1$	8
YIJHOO	DL-Prop-2-ynylglycine	$P2_1/a$	4

carboxylate group. The histogram of intermolecular N(ammonium)–O(carboxylate) distances in the CSD shows a sharp peak at 2.8 Å, which levels off between 3 and 3.2 Å (see supplementary material[2]). Increasing the cutoff value above 3 Å introduces strongly asymmetric bifurcated bonds and additional ammonium–carboxylate contacts which are approximately collinear with the C—N bond. While these interactions may contribute to the stability of the crystal, they are much weaker than normal hydrogen bonds and their inclusion in motif definitions as 'true' hydrogen bonds would reduce the specificity of the analysis (see §4 for an example).

The results obtained with a 3 Å cutoff were manually checked and adjusted for structures in which this cutoff identified two or four hydrogen bonds donated by the same ammonium moiety. Four short N···O contacts were found with the same N atom in GLYCIN16, GLYCIN27, LCYSTN04 and VALIDL02. One of the four contacts could be eliminated in GLYCIN16, GLYCIN27 and VALIDL02 on the basis of the corresponding N—H···O angles being less than 120°. All four contacts were retained for LCYSTN04 [thus including a bifurcation of d(N···O) = 2.89 Å, ∠(N—H···O) = 160° and d(N···O) = 2.96 Å, ∠(N—H···O) = 130° contacts].

In four structures (CEDFAS, FEGMAG, KEPJOE, LCYSTN21) one of the hydrogen bonds is formed with a N···O distance between 3 and 3.11 Å. Originally, only two hydrogen bonds were found in DLABUT11. Voogd & Hulscher (1980) have pointed out that the published coordinates of this structure were wrong, and unfortunately these wrong coordinates are archived in the CSD. After applying the sign corrections suggested by Voogd & Hulscher (1980), the structure of DLABUT11 exhibits three hydrogen bonds with N···O distances below 3 Å. The corrected structure of DLABUT11 was used for further analysis and it will be included in future versions of the CSD.

Table 2
Infinite hydrogen-bond chains in the 52 α-amino acid crystal structures.

Chain motif	Graph set†	Frequency (%)
(···OCCNH···)	$C(5)$	100
(···HNH···O···)	$C_2^1(4)$	75
(···HNH···OCO···)	$C_2^2(6)$	100
(···HNH···O···HNCCO···)	$C_3^2(9)$	90
(···HNH···OCO···HNCCO···)	$C_3^3(11)$	100
(···HNH···O···HNH···OCO···)	$C_4^3(10)$	100
(···HNH···O···HNCCO···HNCCO···)	$C_4^3(14)$	100
(···HNH···OCCNH···O···HNCCO···)	$C_4^3(14)$	100
(···HNH···OCO···HNCCO···HNCCO···)	$C_4^4(16)$	100
(···HNH···OCCNH···OCO···HNCCO···)	$C_4^4(16)$	100

† Since the symmetry equivalence of the hydrogen bonds was ignored, the same motif may be represented by different graph sets in different structures. Only the shortest chain path is given in the table, *i.e.* $C(5)$ or $C_1^1(5)$ also represents $C_2^2(10)$, $C_3^3(15)$ *etc.* graph sets.

3.1. Chain motifs

The hydrogen-bond network of each structure is infinite in at least one dimension, with tape, layer or three-dimensional

[2] Supplementary data for this paper are available from the IUCr electronic archives (Reference: GP5024). Services for accessing these data are described at the back of the journal.

Table 3
Ring motifs formed by 2–5 molecules that were found in at least four structures.

An 'x' marks primitive rings (see text), while a '+' sign denotes examples where the motif is found only as formed by the fusion of smaller rings. An '*' after the REFCODE shows structures with $Z' = 2$.

Motif	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Graph set	$R_2^2(10)$	$R_3^3(11)$	$R_4^2(8)$	$R_4^3(10)$	$R_4^4(12)$	$R_4^3(14)$	$R_4^3(14)$	$R_4^4(16)$	$R_4^4(16)$	$R_4^4(20)$	$R_5^3(13)$	$R_5^4(15)$	$R_5^4(15)$	$R_5^5(17)$	$R_5^4(19)$	$R_5^4(19)$	$R_5^5(21)$	$R_5^5(21)$
Frequency	52%	12%	17%	13%	40%	62%	58%	83%	54%	56%	21%	27%	23%	21%	27%	31%	37%	31%
BOQCUF					x	x	x	x		x								
CERGEL					x	x	x	x		x								
CHEDGL10					x	x	x	x		x								
DLABUT03					x	x	x	x		x								
DLABUT05					x	x	x	x		x								
DLMETA02					x	x	x	x		x								
DLMETA05					x	x	x	x		x								
DLNLUA01					x	x	x	x		x								
DLNLUA02					x	x	x	x		x								
FIVGEW					x	x	x	x		x								
LALXAX					x	x	x	x		x								
LARSUT					x	x	x	x		x								
XADTUR*					x	x	x	x		x								
YIJHOO					x	x	x	x		x								
FEGMAG						+		x								+	+	
LCYSTN21						+		x								+	+	
LCYSTN24						+		x								+	+	
LCYSTN26*						+	x									+	+	
GLYCIN36				x		x			x									
GLYCIN68				x		x			x									
GLYCIN16						x	x	x	x									
GLYCIN27							x	x	x									
ALUCAL04		x					x		+						+		+	+
LALNIN12		x					x		+						+		+	+
DLALNI01		x				x			+							+		+
CEDFAS	x	x	x				+		+	+		+			+		+	+
EXAXOQ*	x	x	x				+		+	+		+		+	+		+	+
SIMRAH	x	x	x							+		+	+	+	+			
VALIDL02	x		x		x		+	+	x	+								
VALIDL	x		x		x		+	+	x	+								
DLILEU02	x		x		x		+	+	x	+								
DLLEUC	x		x		x		+	+	x	+								
GLYCIN19	x		x		x		+	+	x	+								
AMMPRA01	x		x		x			+		x								
DLABUT11	x				x	x		+		+								
ACYHXA01	x			x			+	+		x								
ACXTPY	x			x			+	+		+								
EXAXEG*	x			x			+	+	x	+								
LEUCIN03	x			x			+	+	x	+								
SIMPEJ*	x			x			+	+	x	+								
AHEJEC01*	x					x		+	x		x	+	+	x	+	+	+	+
DAILEU*	x					x		+	x		x	+	+	x	+	+	+	+
DEYFOD*	x					x		+	x		x	+	+	x	+	+	+	+
ICAMOO*	x					x		+	x		x	+	+	x	+	+	+	+
JOXBED*	x					x		+	x		x	+	+	x	+	+	+	+
LCYSTN04*	x					x		+	x		x	+	+	x	+	+	+	+
LISLEU02*	x					x		+	x		x	+	+	x	+	+	+	+
LMETON02*	x					x		+	x		x	+	+	x	+	+	+	+
LVALIN01*	x					x		+	x		x	+	+	x	+	+	+	+
LNLEUC10	x							+	x		x	+	+			+	+	+
LEUCIN01*	x							+	+		x	+	+			+	+	+
KEPJOE	x																	

network topology. Consequently, isolated discrete motifs are not found in any of the 52 structures.[3] Infinite hydrogen-bonded chains, on the other hand, are present in each structure (Table 2), and all possible chain motifs are exhibited by more than half of the structures. The least frequent motif, the (···HNH···O···) chain, is not present in structures that display $R_4^2(8)$ rings [(3) in Fig. 2, see below] or $R_3^2(9)$ rings (···HNH···O···HNCCO···). Similarly, the $C_3^2(9)$ chain was not found in most of the structures displaying $R_4^2(8)$ rings.

3.2. Ring motifs

Ring motifs formed by two to six molecules were sought. Two- to five-molecule ring motifs found in at least four

[3] Discrete **D** motifs, according to the definition of Etter (1990), can be found in structures with $Z' > 1$, but only if a single hydrogen bond is considered which links two symmetry-independent molecules.

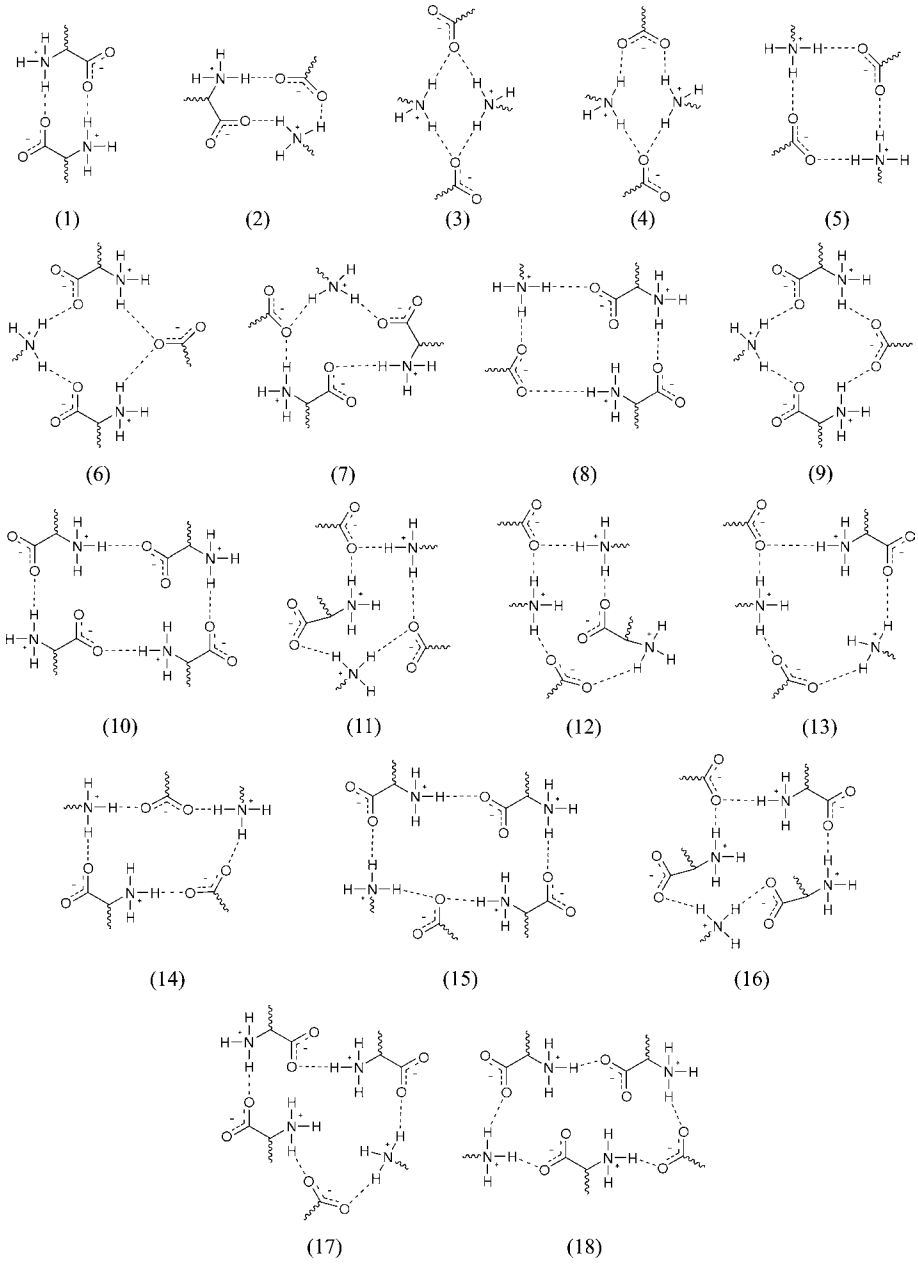

Figure 2
Ring motifs found in at least four of the α-amino acid structures.

structures are shown in Fig. 2, while the structures displaying these motifs are given in Table 3.[4] Ring motifs formed by six molecules are listed in Table 4.

The frequency of ring motifs shows a much larger variation than that of the chain motifs. The $R_2^2(6)$ ring ($\cdots$HNH$\cdots$OCO$\cdots$), for example, was found in only one structure (KEPJOE), while the $R_4^4(16)$ ring [(8) in Fig. 2] is present in 43 of the 52 structures. The frequency of ring motifs seems to be related to the number of molecules forming the ring. Two-molecule rings are present in 52%, three-molecule rings in 21%, four-molecule rings in 98%, five-molecule rings in 40% and six-molecule rings in 100% of the structures. Larger rings and those formed by an even number of molecules thus appear to be more frequent.

The higher frequency of larger rings is not surprising. Larger rings are generally more flexible, thus imposing less restriction on the mutual arrangement of the molecules that form the ring. This, in turn, makes it less likely that an otherwise favourable motif cannot form due to conflicts with close packing.

Larger rings may also be formed by the fusion of smaller rings. The smallest such ring in the structures surveyed is the $R_4^3(14)$ motif [(7) in Fig. 3]. This motif is generated by the fusion of two smaller ring motifs, (1) and (4) in ACYHXA01 (Fig. 3*a*). By contrast, there are no smaller hydrogen-bonded rings within the same $R_4^3(14)$ motif in FIVGEW (Fig. 3*b*), although it is stabilized by a CH$\cdots$O interaction. In graph theory, cycles that contain no chord edges, which would split them into smaller cycles, are called primitive cycles. Analogously, we use the terms primitive (FIVGEW) and non-primitive (ACYHXA01) hydrogen-bond ring motifs. Even though the listing of non-primitive rings may seem redundant, they are included in our analysis because they provide information on the spatial relationships of smaller primitive rings.

The smaller overall frequency of three- and five-molecule rings is presumably related to the crystal symmetry of the amino acid structures (mostly monoclinic and orthorhombic). The formation of five-molecule rings is apparently correlated with the occurrence of $Z'>1$ structures (marked with an asterisk in Table 3). Of the 15 structures with $Z'=2$, 12 contain five-molecule ring motifs, and of the 21 structures with five-molecule rings, only nine crystallize with $Z'=1$. The formation of five-molecule rings is common in Söhncke (or proper) space groups. There are 23 chiral structures in the dataset, and 18 of them pack with five-molecule rings.

Despite the above overall trend, both rare and frequent ring motifs can be formed by the same number of molecules. For example, the six-molecule ring ($\cdots$HNH$\cdots$OCO$\cdots$HNH$\cdots$OCO$\cdots$HNH$\cdots$OCO$\cdots$) is not present in any of the 52 α-

[4] The ring motifs omitted from the table are: $R_3^2(9)$ in FEGMAG, LCYSTN21, LCYSTIN24 and LCYSTIN26, $R_2^2(6)$ in KEPJOE, $R_3^3(15)$ in SIMRAH and $R_5^5(25)$ in EXAXOQ. The first two motifs are discussed in the text, while the last two motifs are formed by three and five HNCCO fragments, respectively.

Table 4
Frequency of ring motifs formed by six molecules in the α-amino acid structures.

Ring motif	Graph set	Frequency (%)
(···HNH···O···HNH···O···HNH···O···)	$R_6^3(12)$	2
(···HNH···O···HNH···OCO···HNH···OCO···)	$R_6^5(16)$	54
(···HNH···O···HNH···OCCNH···O···HNCCO···)	$R_6^4(18)$	54
(···NH···O···HNCCO···HNH···O···HNCCO···)	$R_6^4(18)$	25
(···HNH···O···HNCCO···HNH···OCCNH···O···)	$R_6^4(18)$	15
(···HNH···O···HNH···O···HNCCO···HNCCO···)	$R_6^4(18)$	4
(···HNH···O···HNH···OCO···HNCCO···HNCCO···)	$R_6^5(20)$	73
(···HNH···O···HNCCO···HNH···OCO···HNCCO···)	$R_6^5(20)$	67
(···HNH···O···HNCCO···HNCCO···HNH···OCO···)	$R_6^5(20)$	54
(···HNH···O···HNCCO···HNH···OCCNH···OCO···)	$R_6^5(20)$	33
(···HNH···OCO···HNH···OCCNH···O···HNCCO···)	$R_6^5(20)$	33
(···HNH···O···HNH···OCCNH···OCO···HNCCO···)	$R_6^5(20)$	29
(···HNH···OCO···HNH···OCO···HNCCO···HNCCO···)	$R_6^6(22)$	73
(···HNH···OCO···HNCCO···HNH···OCO···HNCCO···)	$R_6^6(22)$	69
(···HNH···OCO···HNCCO···HNH···OCCNH···OCO···)	$R_6^6(22)$	62
(···HNH···OCO···HNH···OCCNH···OCO···HNCCO···)	$R_6^6(22)$	48
(···HNH···OCCNH···OCCNH···O···HNCCO···HNCCO···)	$R_6^5(24)$	69
(···HNH···OCCNH···O···HNCCO···HNCCO···HNCCO···)	$R_6^5(24)$	65
(···HNH···O···HNCCO···HNCCO···HNCCO···HNCCO···)	$R_6^5(24)$	42
(···HNH···OCCNH···OCO···HNCCO···HNCCO···HNCCO···)	$R_6^6(26)$	85
(···HNH···OCCNH···OCCNH···OCO···HNCCO···HNCCO···)	$R_6^6(26)$	69
(···HNH···OCO···HNCCO···HNCCO···HNCCO···HNCCO···)	$R_6^6(26)$	69
(···OCCNH···OCCNH···OCCNH···OCCNH···OCCNH···OCCNH···)	$R_6^6(30)$	56

amino acid structures, while other six-molecule rings (Table 4) were found in more than half of them. Motifs that are composed of the same molecular fragments in a different sequence have identical graph-set descriptors. Yet, there may be a large difference between the frequencies of such rings. The $R_6^5(20)$ rings in Fig. 4 differ only in the sequence of two molecular fragments, but their frequencies show a twofold difference.

In contrast to six-molecule rings, the frequencies of five-molecule rings vary little. Indeed, Table 3 shows a group of 11 structures that all display the five-molecule ring motifs (11), (12), (13), (16), (17) and (18), thereby accounting for their similar frequencies. Inspection of these structures reveals that all these motifs are formed *via* the assembly of six molecules (Fig. 5). One can identify each motif by following different paths through the hydrogen bonds among the six molecules (Table 5). The occurrence of the six five-molecule ring motifs together thus encodes a particular packing arrangement.

The presence of several motifs in the same assembly of molecules suggests that groups of motifs that are seen frequently together may indicate common packing arrangements. Motif combination can then be used to identify structures with common extended patterns of self-assembly, and to find families of similar structures. This idea is supported by the first 14 structures in Table 3, which all display motifs (5), (6), (7), (8) and (10). They all contain hydrogen-bonded bilayers with identical network topologies and similar molecular arrangements. CHEDGL10 and XADTUR, two members of this group, crystallize with different space groups and different Z' values (Table 1). Despite these crystallographic differences and their chemical difference (cyclohexadienyl *versus* allyl substitution), they show remarkably similar packing arrangements (Fig. 6). It is worth noting that these structures are all racemic, while the 11 structures displaying motifs (11), (12), (13), (16), (17) and (18) are all chiral. The two packing

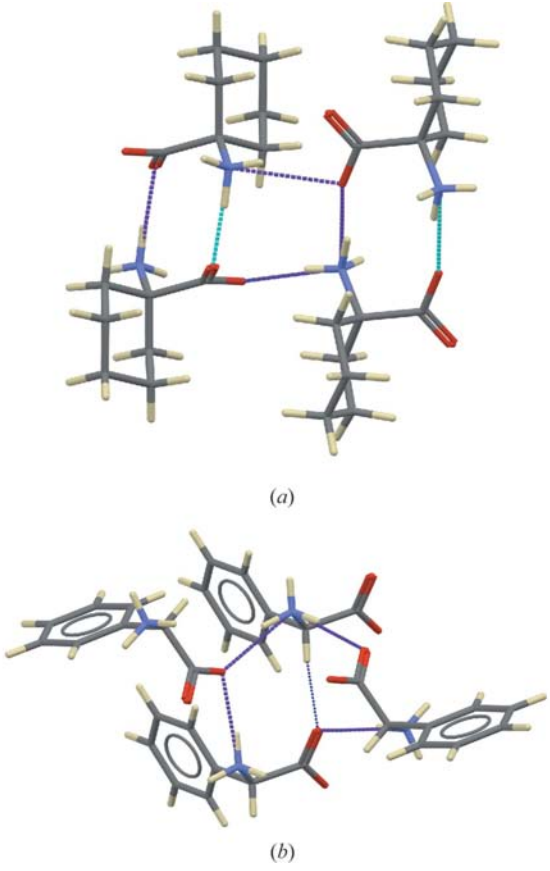

Figure 3
The $R_4^3(14)$ ring motif (7) as found in (*a*) ACYHXA01 and (*b*) FIVGEW. The hydrogen bonds forming the ring motif are shown in purple, other hydrogen bonds in cyan, and a short CH···O contact in blue.

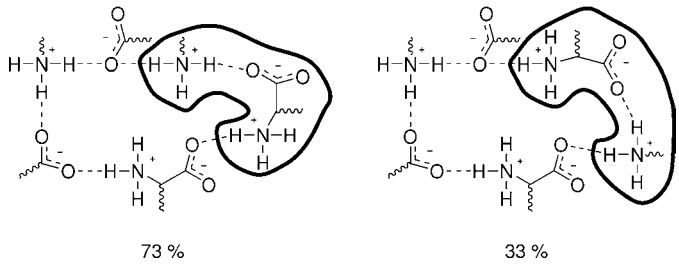

Figure 4
The frequencies of these two six-molecule $R_6^5(20)$ rings are markedly different, even though they differ only in the sequence of the two encircled fragments.

Table 5
Five-molecule ring motifs that can be identified in the same assembly of six molecules (Fig. 5) by following different hydrogen bond paths.

Ring motif†	Graph set	Hydrogen-bond path‡
(11)	$R_5^3(13)$	e–b–f–g–c
(12)	$R_5^4(15)$	c–g–f–a–e
		e–b–h–i–c
(13)	$R_5^4(15)$	e–b–h–i–d
(16)	$R_5^4(19)$	e–b–f–g–d
(17)	$R_5^5(21)$	e–a–h–i–d
		g–d–e–a–f
(18)	$R_5^5(21)$	a–h–i–c–e

† Motifs are defined in Fig. 2. ‡ The paths are given by using the hydrogen-bond labels shown in Fig. 5. The hydrogen bond corresponding to that at the top of the motif diagram in Fig. 2 is given first.

arrangements discussed above thus represent the most frequent patterns of self-assembly of racemic and chiral α-amino acids, respectively.

Analogous comparisons can be made for other groups of structures. The similar hydrogen-bond networks of hydrophobic amino acids with branched chains have been discussed by Dalhus & Görbitz (2000). Their similarity can be seen easily from Table 3 (VALIDL02, VALIDL, DLILEU02 and DLLEUC). While DLILEU02 and DLLEUC are isostructural, the polymorphs VALIDL and VALIDL02 must differ. Both forms are built from hydrogen-bonded bilayers with the same network topology and same arrangement of the molecules. They differ only in the stacking of the bilayers. There are no hydrogen bonds between the bilayers, only van der Waals interactions. We can thus conclude that combinations of motifs that occur together identify extended hydrogen-bond networks, which in turn, enforce a particular arrangement of the molecules. This arrangement may determine the entire crystal structure or a well defined part of it.

3.3. Comparison of polymorphs

Assessing the stability and kinetic accessibility of polymorphs is of great importance in the development of pharmaceutical formulations, and has attracted significant scientific interest in the context of crystal engineering and crystal structure prediction. Reference to already known crystal structures may hint whether a predicted polymorph or an actual structure is likely to form and be stable. We have demonstrated in the previous section that listings of hydrogen-bond motifs provide a quick way to compare structures. Since our dataset contains polymorphs of six compounds, the use of such comparisons in the assessment of polymorphs has been briefly explored.

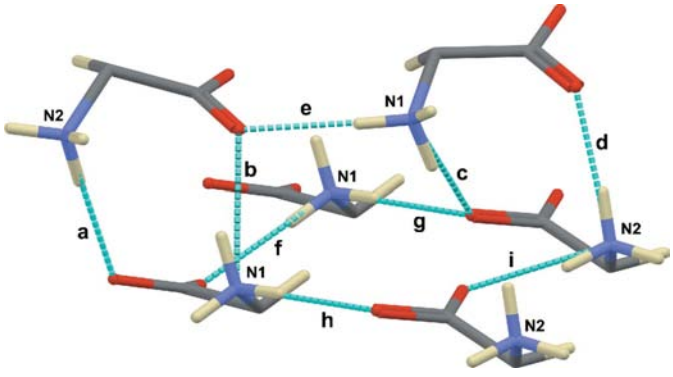

Figure 5
Five-molecule rings in LVALIN01 are formed by two dimers [both $R_2^2(10)$, hydrogen bonds (a, b) and (c, d)] that are linked *via* either the ammonium (f, g) or the carboxylate (h, i) group of a fifth molecule. The isopropyl groups were omitted for clarity. N atoms are labelled to distinguish the two symmetry-independent molecules.

As discussed above, the DL-valine polymorphs VALIDL and VALIDL02 exhibit the same hydrogen-bond motif, so any difference in their stability must be related to weaker interactions, not hydrogen bonding. The same argument applies to the polymorphs of DL-methionine (DLMETA02 and DLMETA05) and DL-norleucine (DLNLUA01 and DLNLUA02).

There are two ambient and two high-pressure polymorphs of L-cysteine in the α-amino acid dataset. The NH$\cdots$O hydrogen-bond motifs are the same in form (I) (LCYSTN21) and in both high-pressure forms (LCYSTN24 and LCYSTN26). These forms differ in the conformation of the cysteine molecules and in the interactions of the thiol group (see Moggach *et al.*, 2005, 2006, for a detailed discussion). Form (I) (LCYSTN21) and form (II) (LCYSTN04) also show some similarity. The ring motifs (6), (8), (16) and (17) are present in both forms. In addition to the common motifs, an $R_3^2(9)$ ring ($\cdots$HNH$\cdots$O$\cdots$HNCCO$\cdots$) was found in LCYSTN21. This motif (not shown in Table 3) is present only in the isostructural pair LCYSTN21 and FEGMAG, and in the high-pressure forms LCYSTN24 and LCYSTN26. The other form, LCYSTN04, includes two further primitive rings: (1) and (11). Both of these rings are common. In addition, exactly the same set of two- to five-molecule rings is found in LCYSTN04 as in eight other structures. All these observations favour LCYSTN04 as the 'most likely' L-cysteine structure. It remains to be seen if the more usual hydrogen-bond network translates to thermodynamic (or kinetic) stability of this form.

Five forms of glycine were among the 52 structures analysed (Table 1). Two of them (GLYCIN36 and GLYCIN68) are high-pressure forms that are unstable under ambient conditions (Dawson *et al.*, 2005). The thermodynamic stabilities of the other three forms were investigated by Boldyreva *et al.* (2003). They found the γ form (GLYCIN16) to be the most stable and the β form (GLYCIN27) to be the least stable at ambient conditions. At temperatures above 440 K, the α form (GLYCIN19) becomes the most stable polymorph. The α form was found to form more readily than the stable γ form and hardly ever transformed to it. Therefore, GLYCIN16 represents the thermodynamically stable form, while GLYCIN19 is the kinetically most accessible (and a 'kinetically stable') form.

The hydrogen-bonded ring motifs of GLYCIN19 (α form) match those of racemic valine and leucine (VALIDL, VALIDL02 and DLLEUC), while the other forms present unique combinations of motifs (Table 3). A rare motif, the $R_4^3(10)$ ring (4), is present in the high-pressure forms. (All the other molecules that form this motif have aliphatic or aromatic

ring substituents.) In addition, the high-pressure forms lack the frequent $R_4^4(16)$ ring (8), which is present in the three ambient pressure forms. The remaining two polymorphs, GLYCIN16 (γ) and GLYCIN27 (β), show only frequent hydrogen-bond motifs. There is only one motif, the $R_4^3(14)$ ring (6), which is present in the stable form (GLYCIN16) but not in the metastable form (GLYCIN27). This motif was also found in the racemic alanine structure DLALNI01.

The analysis of hydrogen-bond motifs thus favours GLYCIN19, which has the same motifs as four other structures. This form is easily obtained and kinetically stable according to Boldyreva *et al.* (2003), but it is not the thermodynamically stable form. It is easy to conclude from the motif listings that the high-pressure forms GLYCIN36 and GLYCIN68 are the least stable. The difference between the stable GLYCIN16 and the metastable GLYCIN27 is less conclusive, although its similarity with DLALNI01 gives a slight hint to the relative stability of the former structure.

The high-pressure form of L-leucine (LEUCIN03), similar to those of glycine, exhibits the relatively rare motif (3). Interestingly, the same motifs were found in two other structures with $Z' = 2$ as in LEUCIN03 with $Z' = 1$. The ambient-pressure form (LEUCIN01) is formed with motifs that are common to most chiral structures.

Three forms of DL-aminobutyric acid were retrieved from the CSD. The thermodynamic relationship of forms A (DLABUT05) and B (DLABUT11) was analysed by means of solubility measurements (Abraham *et al.*, 1977). While form A was described as 'the common form', form B was found to be more stable at temperatures below *ca* 326 K. The solubility measurements yield a free-energy difference of 0.46 kJ mol^{-1} between the two forms at room temperature. The third, high-temperature form D (DLABUT03) was prepared by heating form A at 353 K for 2 d, and it was stable enough to allow X-ray data collection at room temperature (Nakata *et al.*, 1980).

The same hydrogen-bond motifs were identified in the two less stable forms (Table 3, DLABUT03 and DLABUT05). Several other structures display the same hydrogen-bond network (see *e.g.* Fig. 6), so the occurrence and kinetic stability of both forms is consistent with the frequency of their hydrogen-bond motifs.

Form B (DLABUT11), which is the stable form at room temperature, shows a degree of similarity with the two metastable forms. Motifs (5), (6), (8) and (10) are common to all three forms. In addition to the common motifs, the stable form displays motif (1), while the metastable forms display motif (7). The frequencies of motifs (1) and (7) are similar (52 and 58%, respectively), so motif frequencies provide little discrimination among the three polymorphs. This is in accordance with the small experimental free-energy difference between forms A and B (Abraham *et al.*, 1977), although the frequent hydrogen-bonded bilayer network of forms A and D could falsely suggest that they are more stable. (The stable form B displays a three-dimensional network instead of the bilayers.)

Although many more polymorphic systems need to be investigated to fully assess the utility of motif-based polymorph comparisons, some tentative observations can be made. We found that the polymorphs that display common motifs in common combinations exhibit at least kinetic stability, and are easily obtained experimentally. Unstable forms, on the other hand, were often marked by the presence of rare hydrogen-bond motifs. Differences between the frequencies of individual motifs, especially those of a few per cent, were found to be unreliable indicators of relative thermodynamic stability.

4. Cocrystals of hydrophobic α-amino acids

Cocrystals formed by two amino acids were described in a series of papers by Dalhus & Görbitz (1999*a*,*b*,*c*, 2000). Most

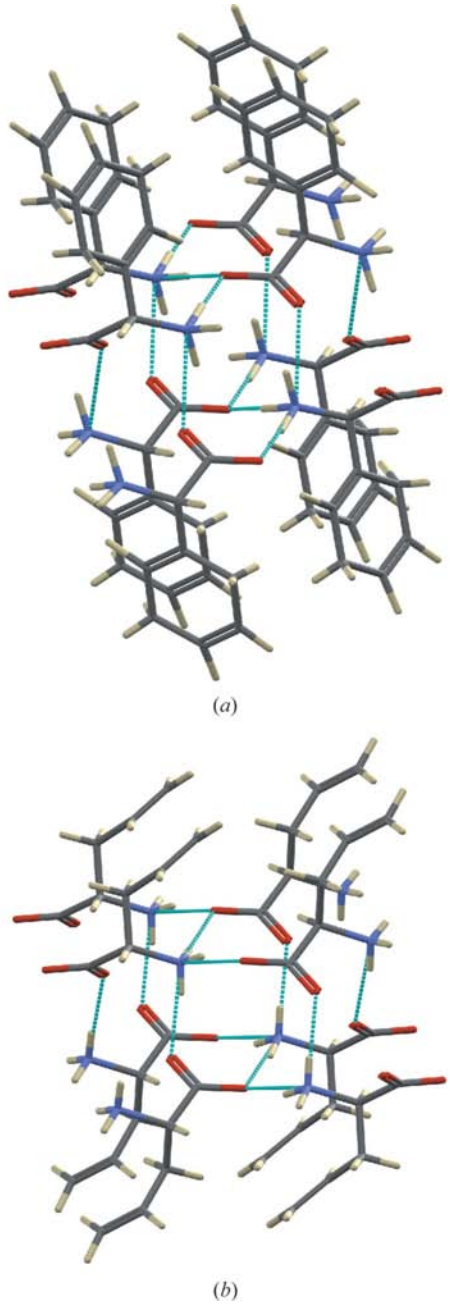

Figure 6
The molecules show similar hydrogen-bonded packing patterns in the crystal structures (*a*) CHEDGL10 and (*b*) XADTUR.

Table 6
Cocrystals of two hydrophobic α-amino acids found in the CSD.

REFCODE	Compounds	Space group	Z
BERNAN	L-Leucine D-2-aminobutanoic acid	$P2_1$	2
BERNER	L-Leucine D-norvaline	$P1$	1
BERNIV	L-Leucine D-methionine	$P2_1$	2
BERPET	L-Leucine D-valine	$P2_1$	2
BERQAQ	L-Valine D-2-aminobutanoic acid	$P2_1$	2
BERQEU	L-Valine D-norvaline	$C2$	4
BERQIY	L-Valine D-methionine	$P2_12_12_1$	4
FITHIZ	L-Isoleucine D-alanine	$P2_1$	2
FITJAT	L-Isoleucine D-2-aminobutanoic acid	$C2$	4
FITJEX	L-Isoleucine D-norvaline	$C2$	4
FITLEZ	l-Isoleucine D-norleucine	$P2_1$	4
FITLID	L-Isoleucine D-methionine	$C2$	4
FITMEA	L-Isoleucine D-valine	$P2_1$	2
FITNIF	L-Isoleucine D-leucine	$P1$	1
GOLVIM	L-Norvaline D-norleucine	$P2_1$	2
GOLVOS	L-Methionine D-norleucine	$P2_1$	2
GOLVUY	L-Valine D-norleucine	$C2$	4
GOLWAF	L-*allo*-Isoleucine D-norleucine	$C2$	4
GOLWEJ	L-Leucine D-norleucine	$P2_1$	2
XADVED	L-Isoleucine D-*allo*-isoleucine	$P1$	1
XADVIH	DL-Isoleucine DL-*allo*-isoleucine	$P\bar{1}$	2

Table 7
Two- to five-molecule ring motifs found in the cocrystals of α-amino acids.

The numbers refer to motifs shown in Fig. 2.

REFCODEs	Motifs
BERNAN, BERNER, BERNIV,† BERPET, BERQAQ, BERQEU, BERQIY, FITHIZ, FITJAT, FITJEX, FITLEZ, FITLID, FITNIF, GOLVIM, GOLVOS, GOLVUY, GOLWAF, GOLWEJ	(5), (6), (7), (8), (10)
FITMEA, XADVED, XADVIH	(1), (3), (5), (7), (8), (9), (10)

† This structure displays the following further motifs if both N···O contacts of a bifurcated bond are considered (see text): (2), (11), (12), (13), (14), (15), (17), (18), $R_3^3(15)$, $R_5^5(25)$, $R_3^2(9)$.

of these cocrystals (Table 6) are quasi-racemates, *i.e.* they are formed by pairs of sterically similar D- and L-amino acids. Ammonium groups donate three hydrogen bonds with an N···O distance of less than 3 Å and a hydrogen-bond angle above 120° in each structure except BERNIV. An asymmetric bifurcated hydrogen bond exists in BERNIV [d(N···O) = 2.91 Å, ∠(N—H···O) = 149° and d(N···O) = 2.96 Å, ∠(N—H···O) = 121°], which will be used to illustrate the effect of including such bifurcations in the definition of the motifs.

The ring motifs found in the cocrystal structures are given in Table 7. Most of the structures displays motifs (5), (6), (7), (8) and (10). This is the most frequent set of motifs for racemic α-amino acid structures (Table 3) and corresponds to a hydrogen-bonded bilayer arrangement (Fig. 6). Motifs (1), (3), (5), (7), (8), (9) and (10) were identified in the remaining three structures. These motifs are the same as those found in racemic crystals of three amino acids with branched alkyl substituents (DLILEU02, DLLEUC, VALIDL and VALIDL02) and in GLYCIN19. These results show that the quasi-racemic cocrystals exhibit the same motifs as racemic homomolecular crystals. Furthermore, the most frequent motif combination is the same for both the cocrystals and the homomolecular crystals.

The above classification applies to BERNIV if only the shorter of the bifurcated contacts is included in motif definitions. If both contacts are considered then the number of motifs found in the structure nearly triples (Table 7). This would imply that almost all of the possible motifs are present in BERNIV, some of which are unique to this structure. BERNIV, however, is not a unique structure. It is isostructural with other cocrystals in the space group $P2_1$ (*e.g.* BERNAN and BERPET). The similarity of these structures can be revealed by hydrogen-bond motifs only if a comparable set of hydrogen bonds is considered. In an automatic survey it may be difficult to find a single cut-off value that applies equally to each structure. As illustrated by BERNIV, the presence of many, unusual motifs in a structure may be an indication that the cut-offs used are not appropriate for that particular structure.

Table 8
β- and γ-amino acid structures from the CSD without hydrogen-bonding groups other than the ammonium and carboxylate groups.

REFCODE	Name	Space group	Z
β-amino acids			
BALNIN01	β-Alanine	*Pbca*	8
GIKNOD	(2*S*,3*R*)-3-Amino-2-phenylthiobutanoic acid	$P2_12_12_1$	4
DASBAB	*cis*-2-Aminocyclopentanecarboxylic acid	$P\bar{1}$	2
DASBEF	*cis*-2-Aminocyclohexanecarboxylic acid	$P\bar{1}$	2
DASBIJ	*cis*-2-Aminocycloheptanecarboxylic acid	$P\bar{1}$	2
DASBOP	*cis*-2-Aminocyclooctanecarboxylic acid	$P\bar{1}$	2
DASBUV	*cis*-2-Aminocyclohex-4-enecarboxylic acid	$P2_1/c$	4
DASCAC	3-*exo*-Aminobicyclo(2.2.1)heptane-2-*exo*-carboxylic acid	*Pbcn*	8
DATMAN	*trans*-2-Aminocyclohexanecarboxylic acid	$P2_1/c$	4
DATMER	*trans*-2-Aminocyclohex-4-enecarboxylic acid	$P2_1/c$	4
γ-amino acids			
ACIJEC	Vigabatrin	*Fdd*2	16
CIDDEZ	(*S*)-3-Aminomethyl-5-methylhexanoate	$P2_12_12_1$	4
GAMBUT02	γ-Aminobutyric acid (monoclinic form)	$P2_1/a$	4
GAMBUT04	γ-Aminobutyric acid (tetragonal form)	$I4_1cd$	16
HAXKEW	4-Amino-3-(2-thienyl)butyric acid	$P2_1/c$	4
POSFAE	4-Amino-3-(2-benzothiophen)butyric acid	$P2_1/c$	4
QIMKIG	Gabapentin	$P2_1/c$	4
VICGOE	3-Aminoadamantane-1-carboxylic acid	$P2_1/c$	4

5. β- and γ-Amino acids

A survey of β- and γ-amino acid structures was performed to investigate any further generality of the motifs observed in α-amino acids. The same criteria were used to search the CSD for β- and γ-amino acids as for α-amino acids (see §3). Ten β- and eight γ-amino acid structures were found (Table 8). This number of structures is clearly insufficient for any statistical analysis, so the motifs found in these structures will be discussed in terms of analogous motifs in the α-amino acid survey. Owing to the longer chain between the two functional groups, some motifs differ slightly from those defined for the α-amino acids. Thus, the $R_2^2(10)$ motif [(1) in Fig. 2], for example, expands to an $R_2^2(12)$ ring (···HNCCCO···

HNCCCO$\cdots$) in β-amino acids, and to an $R_2^2(14)$ ring ($\cdots$HNCCCCO$\cdots$HNCCCCO$\cdots$) in γ-amino acids.

Similar to the α-amino acids, almost all possible chain motifs can be identified in each structure. The least frequent chain is again the ($\cdots$HNH$\cdots$O$\cdots$) chain, which was not found in four structures. Three of these structures display the ring motif (3), while there is an intramolecular $S(6)$ ring in the fourth structure.

The listing of ring motifs (Table 9) immediately reveals that HAXKEW and POSFAE form the same hydrogen-bond network as 14 racemic α-amino acid structures and 18 quasi-racemate cocrystals (Tables 3 and 7). The common hydrogen-bond network is illustrated in Figs. 6 and 7.

The hydrogen-bond patterns of eight β-amino acids have been described in detail by Fábián *et al.* (2005). The similarities among seven of these eight structures, formed by closely related molecules, are apparent from the listing of ring motifs (Table 9, rows from DASBAB to DATMER). It is also obvious that DASCAC represents a completely different molecular arrangement (owing to steric reasons).

The common part of the seven similar structures is marked by the presence of ring motifs (1), (5), (8) and (10). The same group of motifs is present in two γ- and seven α-amino acid structures. The common ring motifs encode a ladder motif (Fig. 8), which is present in each structure with such rings. A slightly modified ladder motif (Fig. 3*a*) is indicated by the presence of motifs (1), (4), (7), (8) and (10) in eight structures (Tables 3 and 9).

The recurrence of ring motifs, their combinations and extended hydrogen-bond networks show that similar self-assembly patterns are favourable for α-, β- and γ-amino acids. The results justify the equivalent treatment of 'covalently expanded' motifs. The role of the four-molecule $R_4^3(14)$ ring motif (6) in CHEDGL10 (Fig. 6) is exactly the same as that of the corresponding $R_4^3(18)$ rings in POSFAE (Fig. 7).

The monoclinic form of γ-aminobutyric acid (GAMBUT02) displays the common ladder motif [*i.e.* (1), (5), (8) and (10)], while the tetragonal form (GAMBUT04) exhibits two of the less common ring motifs (Table 9). This suggests that the monoclinic form is more stable. Although we could not find data on the thermodynamic relationship of these polymorphs, the stability of the monoclinic form is backed by the number of redeterminations archived in the CSD. The structure of the monoclinic form has been reported three times, while that of the tetragonal form only once.

Table 9
Two- to five-molecule ring motifs found in the β- and γ-amino acid structures.

The numbers refer to motifs of α-amino acids shown in Fig. 2, but some of those motifs are expanded in these structures to account for the longer covalent link between the two functional groups (see text).

REFCODE	Motifs
ACIJEC	(2), (6), (9), (16), (18)
BALNIN01	(1), (4), (7), (8), (10), (9)
CIDDEZ	(1), (4), (7), (8), (10)
GIKNOD	(8)
DASBAB	(1), (5), (8), (10)
DASBEF	(1), (5), (8), (10)
DASBIJ	(1), (5), (8), (10)
DASBOP	(1), (5), (8), (10)
DASBUV	(1), (5), (8), (10)
DATMAN	(1), (5), (8), (10), (3), (7)
DATMER	(1), (5), (8), (10), (3), (7)
DASCAC	$R_3^2(10)$†, (2), (6), (8)
GAMBUT02	(1), (5), (8), (10), (6)
GAMBUT04	$R_3^2(11)$,† (18)
HAXKEW	(5), (6), (7), (8), (10)
POSFAE	(5), (6), (7), (8), (10)
QIMKIG	(1), (5), (8), (10), (3)
VICGOE	(1), (4), (7), (8), (10)

† This motif is analogous to the $R_3^2(9)$ rings found in two α-amino acid structures, FEGMAG and LCYSTN21.

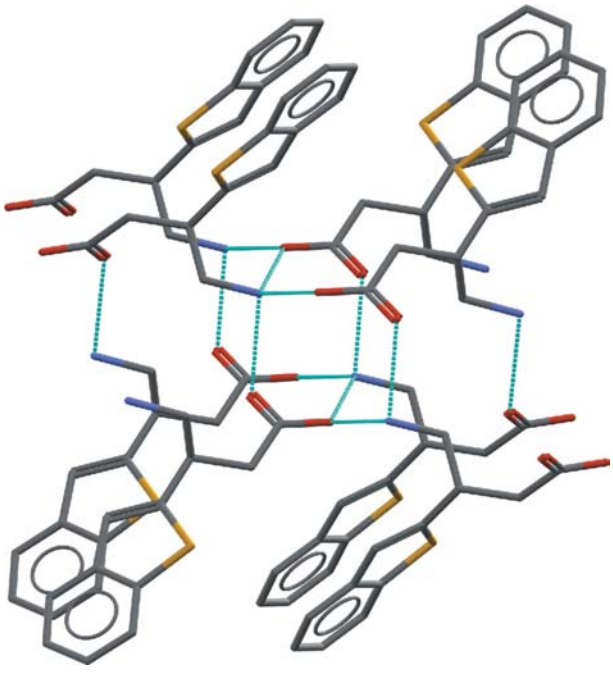

Figure 7
The hydrogen-bond network in the γ-amino acid structure POSFAE is essentially the same as in the α-amino acid structures CHEDGL10 and XADTUR.

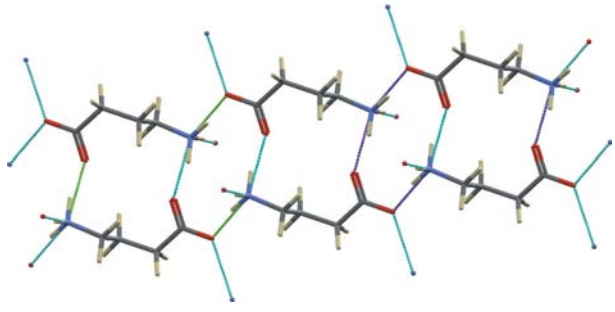

Figure 8
Ladder motif in GAMBUT02. The ladder motif is formed by alternating ring motifs (1) and (5). Motif (8) is indicated by purple hydrogen bonds, while motif (10) is closed by three green bonds and the leftmost purple one.

6. Conclusion

A new method has been presented to survey hydrogen-bond motifs in a set of crystal structures and it has been used to describe the hydrogen-bond networks of simple amino acid crystals. The motif survey results reveal which structural features are common and which are unique. Sets of motifs that occur together in several structures imply the presence of common extended hydrogen-bond networks in these structures. The common networks were found to result in common packing arrangements, thereby leading to the recognition of structural classes among the amino acid crystals. The motifs and motif combinations that are frequent among α-amino acids recur in β- and γ-amino acid structures and in amino acid cocrystals.

The initial results show that motif frequencies and structural classification may be useful in comparing and assessing polymorphs. The high-pressure forms of glycine, for example, display rare motifs, while the most easily obtained and kinetically stable α-form belongs to a popular structural class. Such an analysis could be performed on computationally predicted polymorphs as well, and guide the selection of the most likely structures from predictions with nearly equal energies.

Pfizer Inc is gratefully acknowledged for funding *via* the Pfizer Institute for Pharmaceutical Materials Science. We thank Frank Allen (CCDC) and William Jones (University of Cambridge) for valuable discussions.

References

Abraham, M. H., Ah-Sing, E., Marks, R. E., Schulz, R. A. & Stace, B. C. (1977). *J. Chem. Soc. Faraday Trans. 1*, **73**, 181–185.
Adsmond, D. A. & Grant, D. J. W. (2001). *J. Pharm. Sci.* **90**, 2058–2077.
Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Bernstein, J., Davis, R. E., Shimoni, L. & Chang, N.-L. (1995). *Angew. Chem. Int. Ed. Engl.* **34**, 1555–1573.
Blagden, N. & Davey, R. J. (2003). *Cryst. Growth Des.* **3**, 873–885.
Boldyreva, E. V., Drebushchak, V. A., Drebushchak, T. N., Paukov, I. E., Kovalevskaya, Y. A. & Shutova, E. S. (2003). *J. Therm. Anal. Calorim.* **73**, 409–418.
Chisholm, J. A. & Motherwell, S. (2004). *J. Appl. Cryst.* **37**, 331–334.
Cruz Cabeza, A. J., Day, G. M., Motherwell, W. D. S. & Jones, W. (2006). *J. Am. Chem. Soc.* **128**, 14466–14467.
Dalhus, B. & Görbitz, C. H. (1999*a*). *Acta Cryst.* C**55**, 1105–1112.
Dalhus, B. & Görbitz, C. H. (1999*b*). *Acta Cryst.* C**55**, 1547–1555.
Dalhus, B. & Görbitz, C. H. (1999*c*). *Acta Cryst.* B**55**, 424–431.
Dalhus, B. & Görbitz, C. H. (2000). *Acta Cryst.* B**56**, 720–727.
Dawson, R., Allan, D. R., Belmonte, S. A., Clark, S. J., David, W. I. F., McGregor, P. A., Parsons, S. A., Pulham, C. R. & Sawyer, L. (2005). *Cryst. Growth Des.* **5**, 1415–1427.
Etter, M. C. (1990). *Acc. Chem. Res.* **23**, 120–126.
Fábián, L. & Kálmán, A. (2004). *Acta Cryst.* B**60**, 547–558.
Fábián, L., Kálmán, A., Argay, Gy., Bernáth, G. & Gyarmati, Zs. Cs. (2005). *Cryst. Growth Des.* **5**, 773–782.
Gelbrich, T. & Hursthouse, B. (2005). *CrystEngComm*, **7**, 324–336.
Grell, J., Bernstein, J. & Tinhofer, G. (1999). *Acta Cryst.* B**55**, 1030–1043.
Haynes, D. A., Chisholm, J. A., Jones, W. & Motherwell, W. D. S. (2004). *CrystEngComm*, **6**, 584–588.
Jeffrey, G. A. (1997). *An Introduction to Hydrogen Bonding.* New York: Oxford University Press.
Macrae, C. F., Bruno, I. J., Chisholm, J. A., Edgington, P. R., McCabe, P., Pidcock, E., Rodriguez-Monge, L., Taylor, R., van de Streek, J. & Wood, P. A. (2008). *J. Appl. Cryst.* **41**, 466–470.
Moggach, S. A., Allan, D. R., Clark, S. J., Gutmann, M. J., Parsons, S., Pulham, C. R. & Sawyer, L. (2006). *Acta Cryst.* B**62**, 296–309.
Moggach, S. A., Clark, S. J. & Parsons, S. (2005). *Acta Cryst.* E**61**, o2739–o2742.
Motherwell, W. D. S., Shields, G. P. & Allen, F. H. (1999). *Acta Cryst.* B**55**, 1044–1056.
Motherwell, W. D. S., Shields, G. P. & Allen, F. H. (2000*a*). *Acta Cryst.* B**56**, 857–871.
Motherwell, W. D. S., Shields, G. P. & Allen, F. H. (2000*b*). *Acta Cryst.* B**56**, 466–473.
Nakata, K., Takaki, Y. & Sakurai, K. (1980). *Acta Cryst.* B**36**, 504–506.
Nowell, H. & Price, S. L. (2005). *Acta Cryst.* B**61**, 558–568.
Vishweshwar, P., Nangia, A. & Lynch, V. M. (2003). *Cryst. Growth Des.* **3**, 783–790.
Voogd, J. & Hulscher, J. B. (1980). *Acta Cryst.* B**36**, 3178–3179.