Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

A solution to the observed $Z' = 2$ preference in the crystal structures of hydrophobic amino acids

Carl Henrik Görbitz,[a]* Kristian Vestli[a] and Roberto Orlando[b]

[a]Department of Chemistry, University of Oslo, Norway, and [b]Dipartimento di Scienze e Tecnologie Avanzate, Universita' del Piemonte Orientale, Italy

Correspondence e-mail: c.h.gorbitz@kjemi.uio.no

Received 30 January 2009
Accepted 20 April 2009

Chiral amino acids without functional groups in their side chains (hydrophobic amino acids) systematically form crystals with two molecules in the asymmetric unit. In contrast, racemates of the same compounds form crystals with $Z' = 1$. The present investigation addresses the origin of this important difference between enantiomeric and racemic crystals. Through a series of *ab initio* calculations on infinite two-dimensional slabs, derived from crystal structures, as well as calculations on full crystal structures it is shown that it is indeed possible to explain the observed behaviour. Additionally, the (not unexpected) observation that amino acids usually form racemates in the solid phase rather than undergoing racemic separation upon crystallization is rationalized on the basis of energy calculations.

1. Introduction

The number of molecules in the asymmetric unit, Z', can take on a vast number of values ranging from 1/64 to 32 (Steiner, 2000). The normal situation with $Z' = 1$ prevails for about two thirds of the structures in the Cambridge Structural Database (CSD; Version 5.30 of November 2008; Allen, 2002), while $Z' = 2$ ranks third (after $Z' = \frac{1}{2}$) among both organic and metal-organic crystal structures. Usually, the formation of crystals with $Z' \geq 2$ happens in a fairly random and unpredictable manner, but the set of crystal structures of enantiomeric amino acids with hydrophobic side chains, operationally defined here as all side chains that are not involved in strong hydrogen bonds, is special in systematically producing crystals with $Z' = 2$. Although the crystal structures of hydrophobic amino acids have been intensively studied, also with *ab initio* DFT (density-functional theory) methods (Tulip & Clark, 2005), with careful mapping of the hydrogen-bonding patterns (Fábián *et al.*, 2008), no attempts appear to have been made to explain the $Z' = 2$ preference, which is the focus of the present work.

In order to understand the $Z' = 2$ observation, it is first necessary to consider the general construction of the hydrophobic amino acid crystal structures, which all display alternating hydrophobic and hydrophilic regions, Fig. 1.

A hydrophobic region in turn is actually a bilayer, while a hydrophilic region or layer is composed of two hydrophilic sheets. Hydrogen bonding within a sheet uses two of the amino H atoms, while the third serves as a sheet connector. Five distinct ways of constructing a sheet have been found in crystal structures, Fig. 2.

Only one type of sheet, called LD, incorporates a mixture of L and D enantiomers; other sheets contain amino acids of the same hand. Type 1, called L1 when built from L-amino acids and D1 when built from D-amino acids, has the characteristic

16-membered rings (one has been shaded in Fig. 2) with four hydrogen bonds: two donated to carboxylate *anti* lone pairs and two to *syn* lone pairs. The dimensions and the regularity of the pattern is reflected by the two crystallographic axes generating the pattern, with average lengths and standard deviations 5.20 (0.04) and 5.42 (0.03) Å.

Using L1 as a reference, an L2 sheet is obtained after a 1.2 Å sliding motion in the direction of the 5.20 Å axis (arrow in Fig. 2) for every second row of molecules along the 5.42 Å axis, thus increasing Z' from 1 to 2. The length of the shortest axis is consequently retained, 5.22 (0.07) Å, while the second cell parameter shifts to 9.56 (0.09) Å (length averages with standard deviations). As a result of the translation the main acceptor atom of the *trans* amino N—H is unchanged for one of the molecules (molecule *A*), while for the second (molecule *B*) the acceptor shifts to the *syn* lone pair of the opposing O atom, leaving only a long interaction (average 2.36 Å) to the original O atom in what appears to be a three-center interaction (the average length of the short component is 1.92 Å, while the original hydrogen bond in an L1 sheet has an average length of 1.74 Å) in Fig. 2.

When sliding is carried out not just for every second row of molecules in L1 (to give L2), but for all rows, an L3 sheet is obtained. Molecular connectivity in the last alternative in Fig. 2, L*x*, can be seen as an average between the two molecules in L2.

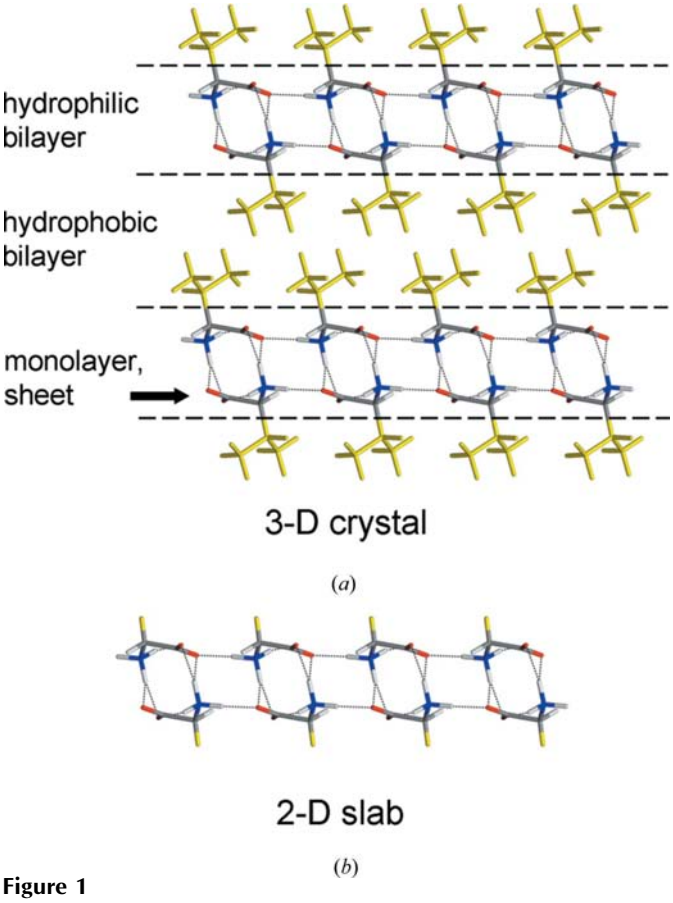

Figure 1
(*a*) General construction of the crystal structure of a hydrophobic amino acid. DL-valine has been used as an example, side chains are depicted in yellow. (*b*) Two-dimensional slab derived from the three-dimensional crystal structure by replacing side chains with H atoms, thus yielding glycine.

Table 1
Hydrophilic layers in amino-acid crystal structures.

Layer	Chirality	Z'	CSD†	Class‡
LD–LD	Racemate	1	32 + 1	I
L1–D1	Racemate	1	17	II
L1–L1	Enantiopure	1§	3	
L2–L2	Enantiopure	2	11	III
L3–L3	Enantiopure	1	1	
L*x*–L*x*	Enantiopure	1	1 + 2	

† Entries in the Cambridge Structural Database (Allen, 2002), monomeric + dimeric ‡ Class definition according to Dalhus & Görbitz (1999*a*). § One structure has $Z' = 2$, but the two molecules have identical hydrogen-bonding environments, while only the side-chain orientations differ.

In putting together two sheets to form a layer, it is imperative that the amino H atoms not involved in hydrogen bonding within the sheets point more or less directly towards carboxylate O-atom acceptors in the adjacent companion sheet. The five sheets shown in Fig. 2 (or their enantiomers) have been found to form layers in six different ways, Table 1.

Four layers, L1–L1, L2–L2 (also called Class III; Dalhus & Görbitz, 1999*a*), L3–L3 and L*x*–L*x*, can be found in structures of enantiomeric compounds. Racemates (or pseudoracemates: 1:1 complexes with one L- and one D-amino acid) have only two different options when forming a layer (and a crystal structure), called LD–LD and L1–D1 (previously Class I and Class II; Dalhus & Görbitz, 1999*a*); the potential L2–D2, L3–D3 and L*x*–D*x* layers have never been found. Accordingly, type 1 sheets (L1 or D1) are unique in being used for the construction of both (pseudo)racemic and enantiopure layers. Previous *ab initio* calculations (Dalhus & Görbitz, 2004) have indicated that the hydrogen-bonding energy of an LD–LD layer is higher than that of an L1–D1 layer, and that the latter will be observed only if the formation of an LD–LD layer is prevented by concomitant steric conflict between hydrophobic side chains.

On a detailed level, an L1–D1 layer is obtained by inversion (inversion centres are present in all structures) of one sheet compared with the next, thus reversing the direction of the N—H···O(*syn*) chains. θ, the angle between similar N—H···O chains in partner sheets, is thus 180°; Fig. 3.

In comparison, enantiopure L1–L1 layers require the presence of a twofold rotation axis, as shown by a red solid line in Fig. 2 (or pseudo-twofold for L-4-fluorophenylalanine; In *et al.*, 2003). The resulting value for θ can be calculated from the cell dimensions, as each chain direction corresponds to a cell diagonal, *e.g.* $\theta = 2\arctan(a/b) = 2\arctan(8.807/5.975\text{ Å}) = 111.7°$ for the high-pressure polymorph of L-leucine (Yamashita *et al.*, 2007).

The two sheets constituting an L2–L2 layer, Fig. 3, are also related by rotation, but with an additional translational element giving a crystallographic 2_1 screw axis. As is evident from the red line in Fig. 2, the axis furthermore runs in a different direction relative to the first sheet compared with L1–L1. The θ angle between is nevertheless calculated as before, *e.g.* $\theta = 2\arctan(a/b) = 2\arctan(9.682/5.247\text{ Å}) = 123.1°$ for L-

Table 2
Crystal structures used in the calculations.

Compound	CSD refcode	Layer/class	Space group	Z'
α-DL-Norleucine[a]	DLNLUA01	LD–LD/I	$P2_1/a$	1
α-Glycine[b]	GLYCIN21	L1–D1/II	$P2_1/n$	1
DL-Valine[c]	VALIDL03	L1–D1/II	$P\bar{1}$	1
DL-Leucine[d]	DLLEUC	L1–D1/II	$P\bar{1}$	1
L-Leucine (hp)[e]†	LEUCIN03	L1–L1	$C2$	1
4-Fluoro-phenylalanine[f]	EXAXEG	L1–L1	$P2_1$	2
L-Valine[g]	LVALIN01	L2–L2/III	$P2_1$	2
L-Leucine[h]	LEUCIN02	L2–L2/III	$P2_1$	2
5-Ethyl L-glutamate[i]	PAZHEE	L2–L2/III	$P2_1$	2
5-Methyl L-glutamate[j]	GAVRAX	L3–L3	$C2$	1
L-Norleucine[k]	LNLEUC10	Lx–Lx	$C2$	1

References: (*a*) Harding *et al.* (1995), (*b*) Langan *et al.* (2002), (*c*) Flaig *et al.* (2002), (*d*) Di Blasio *et al.* (1975), (*e*) Yamashita *et al.* (2007), (*f*) In *et al.* (2003), (*g*) Dalhus & Görbitz (1996), (*h*) Görbitz & Dalhus (1996), (*i*) Wu, Li *et al.* (2005), (*j*) Wu, Xiao *et al.* (2005), (*k*) Torii & Iitaka (1973). † High-pressure polymorph.

valine, Fig. 2 (Dalhus & Görbitz, 1996). θ values for L2–L2 structures fall in a very narrow range between 120.5 and 123.5°.

The L*x* sheet leads to the L*x*–L*x* layer, which, like L1–L1, has $Z' = 1$, but is structurally closely related to L2–L2 with $Z' = 2$. The L3 sheet, on the other hand, which may look quite similar to L*x* in Fig. 2, yields a completely different type of layer in which the two sheets are related by a crystallographic twofold rotation, Fig. 3.

Hydrogen-bonding details in the four enantiomeric layers are shown in Fig. 4.

As indicated above, hydrogen bonding in the L*x*–L*x* layer represents an average of the two molecules in an L2–L2 layer. It is noteworthy that, with the exception of L3–L3, all enantiomeric (Fig. 4) and racemic (not shown) layers include amino acid dimers.

After this basic decoding of the amino acid structures, Table 1 may be studied in more detail. As already indicated, L2–L2 is by far the most common type of enantiomeric layer. The statistics raise two important questions:

(i) Why do hydrophobic amino acids as a rule crystallize with $Z' = 2$ (L2–L2) when patterns with $Z' = 1$ are available?

(ii) If we can show that L2–L2 is the most favourable type of layer, why do we nevertheless occasionally find L1–L1, L3–L3 and L*x*–L*x* structures?

In addition to addressing these fundamental problems concerning the essential building blocks of life, we have also considered the fact that, with the exception of DL-*allo*-isoleucine (Dalhus, 2000), racemic solutions of hydrophobic amino acids do not undergo enantiomeric separation upon crystallization (a property they share with the vast majority of all other organic compounds). This means that, in terms of interaction energies, a supersaturated racemic mixture is better off forming LD crystals than independent L and D crystals. This leads to a third question:

(iii) Can we rationalize why enantiomeric separation does not take place upon crystallization?

Clearly, the observed crystallographic trends must reflect the total interaction energies of the various crystal forms. Answers to the three basic questions were thus sought using *ab initio* calculations on periodic systems with the aid of the computer program *CRYSTAL*06 (Dovesi *et al.*, 2006).

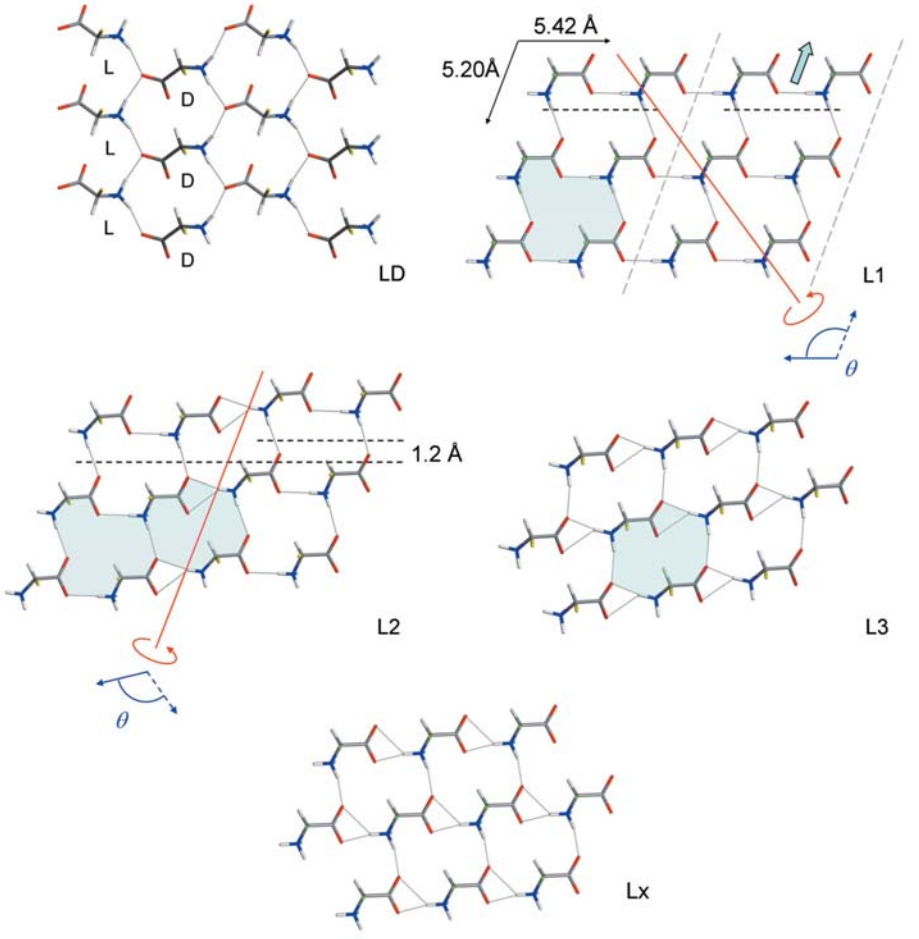

Figure 2
The five observed types of hydrophilic sheets in the crystal structures of hydrophobic amino acids. L and D labels for the LD sheet identifies the two different enantiomers. Side chains have been replaced by H atoms coloured in yellow; characteristic hydrogen-bonded ring systems have been shaded. For an explanation of the dotted and solid lines in red and blue, see text.

2. Methodology

2.1. Crystal structure models

The observation that the formation of L2–L2 structures occurs for a wide variety of side chains with clearly different hydrophobic layers

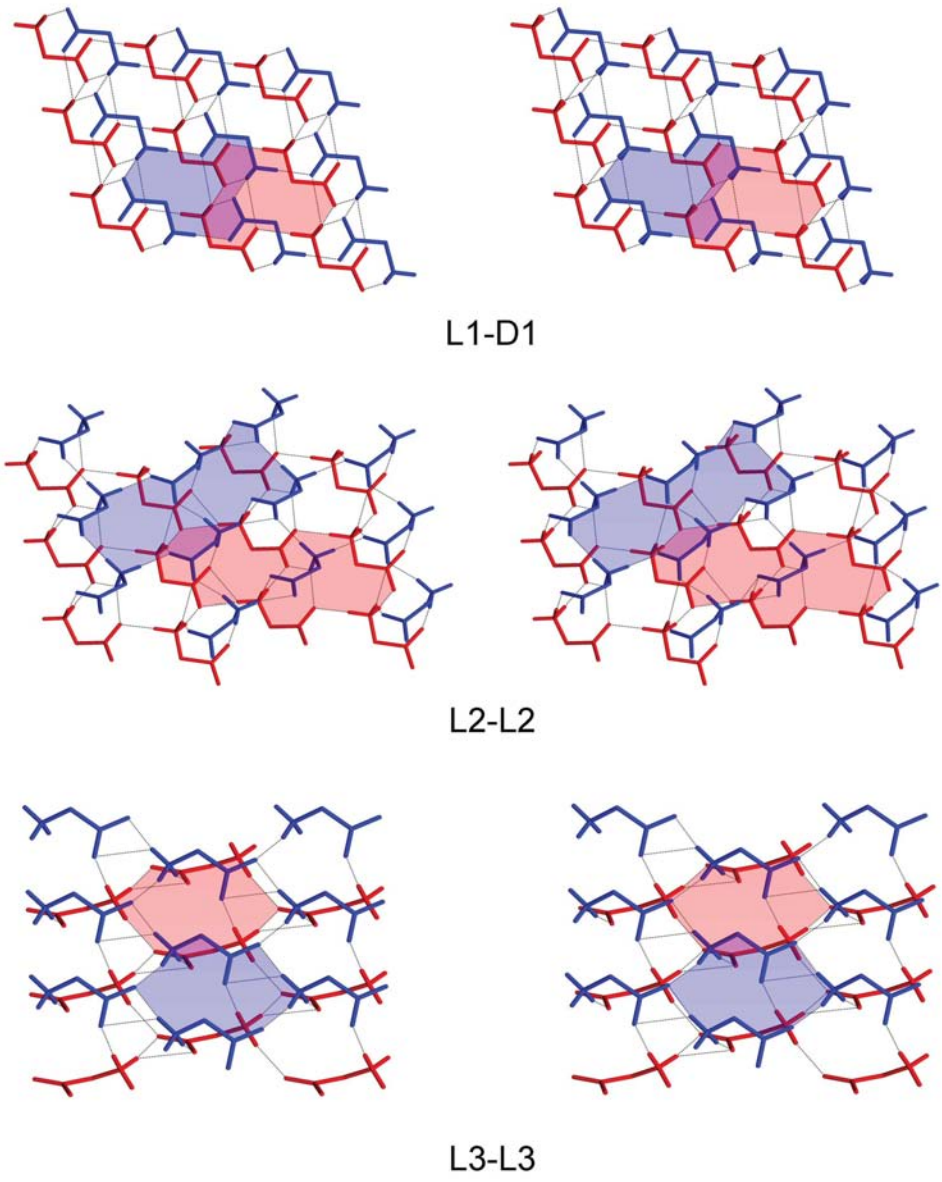

Figure 3
Stereo drawings of individual L1–D1 (top), L2–L2 (middle) and L3–L3 (bottom) layers. The top sheets have been coloured in blue, the bottom sheets in red. Characteristic hydrogen-bonded ring systems within the sheets have been shaded as in Fig. 2. Side chains and C^{α} H atoms have been omitted for clarity.

suggests that the main reason for the $Z' = 2$ preference can be found in the hydrophilic region of the crystal structures and thus is associated with the hydrogen-bonding pattern. To test this hypothesis, 11 high accuracy amino acids were retrieved from the CSD (Allen, 2002), Table 2.

The large L1–D1 and L2–L2 groups are represented by three structures each, L1–L1 by two structures[1] and L3–L3 and L*x*–L*x* by the single available monomeric structure. The LD–LD family is of more peripheral interest to the current investigation and is also represented by just one structure. It is noteworthy that the α polymorph of glycine, despite the obvious lack of side chains that could cause steric conflict, belongs to the L1–D1 group and not LD–LD. 5-Ethyl L-glutamate was included in the L2–L2 set owing to its close molecular similarity with 5-methyl L-glutamate, which forms an L3–L3 structure.

The modelling study started by replacing all side chains with H atoms, thus converting each amino acid to glycine. This procedure was carried out with *Mercury* (Macrae *et al.*, 2006), which was also used to normalize H atom distances to 1.009 Å for N—H and 1.083 Å for C—H.

2.2. *Ab initio* calculations

All calculations were run with a development version of the *CRYSTAL* code in which C. M. Zicovich-Wilson implemented the Grimme empirical model (Civalleri *et al.*, 2008). Despite its very simple form, Grimme's additive atom–atom potential (Grimme, 2004, 2006) adds an estimate of long-range dispersion interactions to the *ab initio* total energy and energy gradients which corrects DFT results fairly effectively. Lattice parameters of amino acids in their crystalline form, which are over-estimated by as much as 10% at the B3LYP level of approximation, are predicted to be within 1–2% of the experimental values after such a correction. Following Civalleri *et al.* we adopted scale factors s_6 = 1.40 for H atoms and s_6 = 1.05 for C, N and O atoms, which appear to be more appropriate for solid-state systems than the original s_6 = 1 used for molecules (Grimme, 2006). The basis sets used for the triple-ζ type + polarization were proposed by Ahlrichs (Schäfer *et al.*, 1992). The use of relatively large split-valence basis sets is recommended in this case in order to reduce the basis-set superposition error as far as possible as it may result in significant overbinding of the systems and the prediction of small unit-cell volumes.

Calculations were performed for two-dimensional bilayers as shown in Fig. 1, utilizing the SLAB command. The correctness of the program input and output files was monitored by the *MOLDRAW* program (Ugliengo *et al.*, 1993).

[1] The third structure in this group (Table 1), D-phenylalanine with CSD refcode SIMPEJ and *R* factor 0.147, was considered too inaccurate to be included.

Table 3
Relative sheet and bilayer energies and interaction energies between sheets in a layer (kJ mol^{-1}) from the optimization of two-dimensional glycine bilayers (side chains replaced by H).

Origin structure	Sheet	E_{rel}	Layer	E_{isheet}	E_{rel}
α-DL-Norleucine	LD	0.0	LD–LD	−41.7	0.0
α-Glycine	L1	15.6	L1–D1	−49.8	7.5
DL-Valine	L1	16.1	L1–D1	−50.3	7.5
DL-Leucine	L1	15.6	L1–D1	−49.8	7.5
L-Leucine (hp)	L1	16.6	L1–L1	−45.8	12.6
4-Fluoro-phenylalanine	L1	16.6	L1–L1	−45.7	12.6
L-Valine	L2	11.7	L2–L2	−47.8	5.6
L-Leucine	L2	11.7	L2–L2	−47.8	5.6
5-Ethyl L-glutamate	L2	11.8	L2–L2	−47.8	5.6
5-Methyl L-glutamate	L3	9.0	L3–L3	−42.5	8.2
L-Norleucine	L*x*	9.4	L*x*–L*x*	−43.3	7.8

Relative energies for various sheets were obtained from single-point calculations based on the optimized layer structures. An estimate of the thermal contribution to the free-energy difference (ΔG) between the two-dimensional glycine layers derived from DL-valine and L-valine at finite T was obtained from the *ab initio* calculation of the phonon vibrational frequencies at the Gamma point only. Phonons for the two structures were calculated within the harmonic approximation in the same computational framework and conditions as for the total energy calculations. ΔG includes the electronic energy at 0 K corrected by the zero-point energy, the thermal contribution to the vibrational energy and the corresponding enthropic contribution. The pressure-volume contribution to ΔH was neglected.

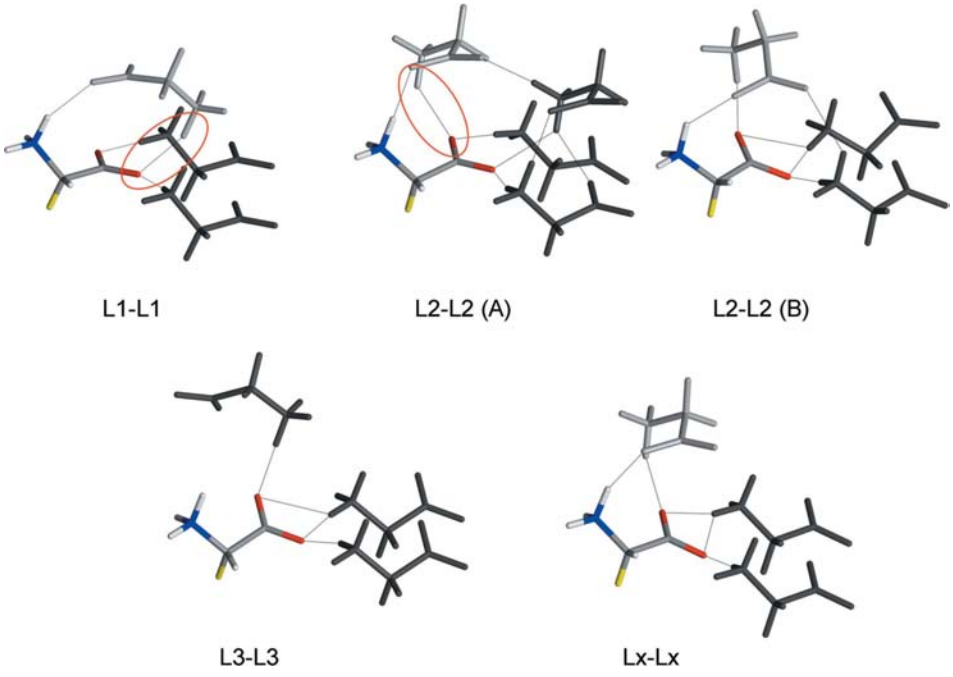

Figure 4
Detailed hydrogen-bonding environments of amino-acid molecules in enantiomeric layers. With the exception of L3–L3, molecules in adjacent sheets form hydrogen-bonded dimers (dimer partner shown in light grey, other amino acids shown in dark grey). The hydrogen bonding in an L*x*–L*x* layer represents an average of the two independent molecules in the L2–L2 layer. Red ellipsoids highlight inter-sheet hydrogen bonds for L1–L1 and L2–L2 layers. This figure is in colour in the electronic version of this paper.

Finally, L-valine, DL-valine, L-leucine and DL-leucine were studied in more detail by full three-dimensional optimization of the crystal structures. Calculation of phonons for the three-dimensional structures was out of reach with the available computer resources.

In all discussions of torsion angles for enantiomers and racemates the conformations given correspond to the L enantiomer.

3. Results and discussion

3.1. Question 1: the Z' = 2 preference: sheets, layers and hydrogen bonds

The calculated relative energies of the observed sheets and layers (obtained from the optimization of glycine bilayers) are listed in Table 3.

Apart from the fact that the LD sheet is lowest in energy, the sequence (from low to high energy) L3/L*x*, L2, L1 is indeed quite unexpected.

Hydrogen bonding between sheets, as reflected by E_{isheet} in Table 3, is however, significantly better for L1–D1 than for LD–LD, reducing the relative energy of the former to 7.5 kJ mol^{-1}. This energy difference is in line with previous *ab initio* calculations (Dalhus & Görbitz, 2004), but slightly smaller owing to a new computation algorithm (previous results were based on single-point calculations for the crystal structures, no optimizations were carried out).

Similarly, for enantiomeric structures L2–L2 hydrogen bonding is more efficient than L*x*–L*x* and L3–L3, resulting in lower layer energy.

The answer to question A is thus that amino acids prefer to crystallize with Z' = 2 (L2–L2) because the resulting hydrogen-bonding pattern is inherently better than the patterns of potential alternative structures with Z' = 1.

3.2. Question 2: when $Z' \neq 2$: the side-chain issue

The choice of a particular hydrogen-bonding sheet introduces specific limitations on the conformational states available to the side-chains. From studies of structure models using the molecular graphics program *SYBYL* (Tripos, 2007), the following properties have been deduced for the various sheets:

(i) L1. An unbranched side chain can in principle take on any

of the three regular conformations for the N—C^{α}—C^{β}—C^{γ} (χ^1) torsion angles, but only *trans* rotamers have been found in crystal structures. Branching at C^{β} (as for valine, isoleucine and *allo*-isoleucine) is permitted, with *trans*/*gauche*+ for $\chi^{1,1}$/$\chi^{1,2}$ as the only viable conformation; model studies suggest that the introduction of a *gauche*+/*gauche*− conformation would inevitably introduce significant steric conflict, the situation becoming even worse for the *trans*/*gauche*− conformation. An L1 sheet can also easily accommodate the C^{γ}-branched side chain of leucine. The *trans*, *trans*/*gauche*+ conformation for χ^1, $\chi^{2,1}$/$\chi^{2,2}$ must then be used and in fact gives a particularly nice close-packing of hydrophobic groups with intermolecular H···H distances (after normalization of C—H distances to 1.083 Å) from 2.53 Å and up. A benzyl side chain with χ^1 = *trans* also fits in easily, as seen in the crystal structure of the L-phenylalanine:D-valine complex (Prasad & Vijayan, 1991).

(ii) L2. Out of the nine basic combinations of conformations for N—C^{α}—C^{β}—C^{γ} (χ^1) torsion angles in the two molecules *A* and *B*, only six are allowed when there is no branching at C^{β} (number of observations in crystals in parentheses): *trans*,*trans* (1), *trans*,/*gauche*+ (1), *gauche*+,*gauche*+ (2), *gauche*−,*trans* (2), *gauche*−,*gauche*+ (0) and *gauche*−,*gauche*−. It follows that branching at C^{β} (as for valine, isoleucine and *allo*-isoleucine) leaves *trans*/*gauche*− for $\chi_A^{1,1}$/$\chi_A^{1,2}$ and *trans*/*gauche*+ for $\chi_B^{1,1}$/$\chi_B^{1,2}$ as the only viable conformation. As for L1, double substitution at C^{β} is not possible; both the L enantiomer and the racemate of *tert*-butylglycine produce quite different hydrated crystal structures (Weissbuch *et al.*, 1990). Any branching at C^{γ} is also disallowed in general, but is tolerated for L-leucine at the expense of substantial deviations from the idealized *trans*,*trans*/*gauche*+ conformation for one of the two molecules in the asymmetric unit (molecule *B*) and uncomfortably short H···H contacts in the range from 2.19 Å and up.

(iii) L3. *Gauche*−, as observed for 5-methyl L-glutamate, and *trans* conformations at χ^1 are allowed. C^{β} branching with $\chi^{1,1}$/$\chi^{1,2}$ set to *trans*/*gauche* may be possible.

(iv) L*x*. All three standard conformations are permitted for χ^1, but branching at C^{β} or C^{γ} is always prohibited.

The paragraphs above have all been associated with steric conflict in hydrophobic regions of the (potential) crystal structures. It is important to realise, however, that a certain side-chain bulk is required for efficient packing; large voids are energetically unfavourable. Accordingly, no structures with enantiomeric sheets incorporating alanine with a simple methyl side chain are known (L-alanine as well as DL-alanine form structures that are not divided into layers). Crystal structures are furthermore missing for L-aminobutyric acid and L-norvaline with ethyl and *n*-propyl side chains. Less obvious is the origin of the observation by Dalhus (2000) that the change in chirality at C^{β} strongly affects the crystal habits for the diastereoisomers isoleucine and *allo*-isoleucine; the latter isomer generally forms lower quality crystals and some pseudoracemic mixtures as well as DL-*allo*-isoleucine itself are separated into enantiomers upon crystallization.

We are then at the point when we can address question B, the observation of L1–L1, L3–L3 and L*x*–L*x* structures, Table 1. It is necessary to consider each of the structures individually.

3.2.1. L-Norleucine. Together with the dimeric amino acids L-cystine (—CH_2—S—S—CH_2— link; Dahaoui *et al.*, 1999) and L-lanthionine (—CH_2—S—CH_2— link; Desiraju & Rao,1990), L-norleucine constitutes a group of L*x*–L*x* structures with unbranched side chains.

The *n*-butyl side chain of L-norleucine may not be able to fill the hydrophobic region of an L2–L2 structure in a satisfactory manner, but the modest substitution of —$C^{\delta}H_2$— with —S— to yield methionine forces a conversion to L2–L2 (Dalhus & Görbitz, 1996), suggesting that the energy preference for L*x*–L*x* is very slim. An L*x*–L*x* layer has a good hydrogen-bonding pattern, but its incompatibility with any side-chain branching means that it is not versatile for general amino acid structures.

3.2.2. Phenylalanine and 4-fluorophenylalanine. In accordance with the model studies, phenylalanine cannot form L2 sheets or L2–L2 layers due to steric conflict, and also not L3–L3 nor L*x*–L*x* layers. It can form L1 sheets, however, and the notorious difficulties obtaining diffraction quality phenylalanine crystals (Khawas, 1970) are associated with problems finding an ordered arrangement at the interface in the centre of the hydrophobic bilayer, Fig. 1. Such a fit can be induced by substitution of a ring H atom in the *para* position with an F atom (4-fluorophenylalanine); *meta* or multiple substitution produces other crystal forms (hydrates or structures without regular sheets; In *et al.*, 2003).

3.2.3. L-Leucine (hp). The excellent packing of leucine side chains for an L1 sheet compared with an L2 sheet suggests that the initial 6.0 kJ mol^{-1} energy difference between glycine-derived L1–L1 and L2–L2 layers in favour of the latter (Table 3) is reduced for the corresponding leucine structures. This is confirmed by three-dimensional *ab initio* optimizations, giving a 3.7 kJ mol^{-1} energy difference in favour of the ambient-pressure L2–L2 polymorph. It is not obvious why high pressure triggers the formation of the L1–L1 polymorph, which at normal pressure and temperature has only marginally higher density than the ambient pressure L2–L2 polymorph, 1.173 and 1.165 g cm^{-1}, respectively. One can only speculate that the high-energy L1–L1 polymorph is more compressible and thus becomes more stable at higher pressures. Once formed, it is also stable at ambient pressure, which is perfectly reasonable as a conversion from L1–L1 to L2–L2 would involve not only breaking strong hydrogen bonds when rotating sheets relative to each other, Fig. 3, but also using a different O atom as an acceptor for the inter-sheet hydrogen bond, Fig. 4. It does not seem very likely that unbranched or C^{β}-branched amino acids could form similar high-pressure polymorphs.

3.2.4. 5-Methyl L-glutamate.. It is hard to give a good explanation for why 5-methyl L-glutamate chooses an L3–L3 structure, but the side chain is very similar to 5-ethyl L-glutamate, which forms an L2–L2 structure. This means that small shifts in the interaction energy at the hydrophobic interface must drive the 5-methyl compound to form an L3–L3 layer. Evidently, these side chains do not pack easily as the slightly distorted L2–L2 layers in the crystal structure involve a

combination of χ^1 torsion angles (molecule *A*: *trans*; molecule *B*: *gauche*−) for the two molecules in the asymmetric unit that would be considered incompatible with such a layer (see above).

3.3. Question 3: enantiomeric separation – putting it all together

The lower energy of the LD–LD layer compared with L2–L2 (and the enantiomer D2–D2) in Table 3 shows that the formation of LD–LD crystals from a racemic solution is clearly energetically favoured over chiral separation into L2–L2 and D2–D2 crystals. If, however, LD–LD crystals cannot be formed due to steric conflict (see above) and the potential racemic crystallization outcome is instead represented by L1–D1, the picture becomes more complex. Table 3 shows that although the inter-sheet interaction energy is higher for L1–D1 than for any other type of layer, it is still not enough to completely cancel out the inherently lower energy of an L2 sheet compared with L1. Accordingly, L1–D1 is marginally higher in energy than L2–L2. The energy difference is 1.9 kJ mol^{-1} in Table 3, while our estimate for ΔG at 298 K (with the inclusion of entropy and thermal energy contributions) is slightly higher, 3.0 kJ mol^{-1}. In itself, this result suggests that enantiomeric separation of such a racemic mixture may take place upon crystallization. As this does not actually happen, the effect of side chains, neglected until now in the simplified glycine model being used, must be considered.

Optimization of the full three-dimensional crystal structures were carried out for L- and DL-valine. The inclusion of side chains tipped the 1.9 kJ mol^{-1} preference for separation (Table 3) for the glycine derivative into a 3.6 kJ mol^{-1} preference for racemate formation, a small but most significant shift. We expect the findings for valine to be fairly representative for hydrophobic amino acids in general, and the calculated preference for leucine racemate formation is quite similar, 3.9 kJ mol^{-1}.

In our calculations we have not considered kinetic effects, which may very well play an important role in determining the crystallization outcome (see *e.g.* Gavezzotti, 2000) or other factors that may be operational, but we find that the observed trends in crystallization of hydrophobic amino acids may be rationalized without invoking such effects. Our results from *ab initio* calculations of thermodynamic stability alone should clearly be treated with some caution, but indicate that the formation of racemic L1–D1 crystals does not take place because of a better hydrogen-bonding arrangement than in chiral crystals, but due to slightly better stacking of hydrophobic side chains. The small energy differences involved do, however, hold open the possibility for finding crystallization conditions (solvent, temperature) that will indeed separate the components of some racemic mixtures. DL-valine could be a candidate; the closely related L-isoleucine:D-valine pseudo-racemates forms a complex in the crystal (Dalhus & Görbitz, 1999*b*), but L-valine:D-*allo*-isoleucine is separated into enantiomers upon crystallization (Dalhus, 2000).

4. Conclusion

We have successfully used two- and three-dimensional *ab initio* optimizations on the solid-state structures of hydrophobic amino acids to show that for enantiopure crystals a hydrogen-bonding pattern with $Z' = 2$ is preferred energetically over alternative arrangements with $Z' = 1$. Compounds with aromatic side chains can, however, not use the inherently more favourable hydrogen-bonding arrangement with $Z' = 2$ owing to inevitable steric conflict. Other occasional observations of structures with $Z' = 1$ can be rationalized on the basis of the need for a certain side-chain bulk in order to avoid unfavourable voids or cavities in the hydrophobic regions of the crystal. Finally, the driving force for the typical outcome of a crystallization of a racemic amino acid solution, *i.e.* racemic crystals (as opposed to enantiomeric separation), may sometimes not be the formation of a better hydrogen-bonding arrangement, but rather a more favourable way of packing the hydrophobic side chains.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Civalleri, B., Zicovich-Wilson, C. M., Valenzano, L. & Ugliengo, P. (2008). *CrystEngComm*, **10**, 405–410.
Dahaoui, S., Pichon-Pesme, V., Howard, J. A. K. & Lecomte, C. (1999). *J. Phys. Chem. A*, **103**, 6240–6250.
Dalhus, B. (2000). PhD thesis. University of Oslo, Norway.
Dalhus, B. & Görbitz, C. H. (1996). *Acta Chem. Scand.* **50**, 544–548.
Dalhus, B. & Görbitz, C. H. (1999*a*). *Acta Cryst.* C**55**, 1105–1112.
Dalhus, B. & Görbitz, C. H. (1999*b*). *Acta Cryst.* B**55**, 424–431.
Dalhus, B. & Görbitz, C. H. (2004). *J. Mol. Struct. Theochem*, **675**, 47–52.
Desiraju, G. R. & Rao, D. R. (1990). *Acta Cryst.* C**46**, 627–629.
Di Blasio, B., Pedone, C. & Sirigu, A. (1975). *Acta Cryst.* B**31**, 601–602.
Dovesi, R., Saunders, V. R., Roetti, C., Orlando, R., Zicovich-Wilson, C. M., Pascale, F., Civalleri, B., Doll, K., Harrison, N. M., Bush, I. J., D'Arco, P. & Llunell, M. (2006). *CRYSTAL*06. University of Turin, Italy.
Fábián, L., Chisholm, J. A., Galek, P. T. A., Motherwell, W. D. S. & Feeder, N. (2008). *Acta Cryst.* B**64**, 504–514.
Flaig, R., Koritsanszky, T., Dittrich, B., Wagner, A. & Luger, P. (2002). *J. Am. Chem. Soc.* **124**, 3407–3417.
Gavezzotti, A. (2000). *J. Am. Chem. Soc.* **122**, 10724–10725.
Görbitz, C. H. & Dalhus, B. (1996). *Acta Cryst.* C**52**, 1754–1756.
Grimme, S. (2004). *J. Comput. Chem.* **25**, 1463–1473.
Grimme, S. (2006). *J. Comput. Chem.* **27**, 1787–1799.
Harding, M. M., Kariuki, B. M., Williams, L. & Anwar, J. (1995). *Acta Cryst.* B**51**, 1059–1062.
In, Y., Kishima, S., Minoura, K., Nose, T., Shimohigashi, Y. & Ishida, T. (2003). *Chem. Pharm. Bull.* **51**, 1258–1263.
Khawas, B. (1970). *Acta Cryst.* B**26**, 1919–1922.
Langan, P., Mason, S. A., Myles, D. & Schoenborn, B. P. (2002). *Acta Cryst.* B**58**, 728–733.
Macrae, C. F., Edgington, P. R., McCabe, P., Pidcock, E., Shields, G. P., Taylor, R., Towler, M. & van de Streek, J. (2006). *J. Appl. Cryst.* **39**, 453–457.
Prasad, G. S. & Vijayan, M. (1991). *Acta Cryst.* C**47**, 2603–2606.
Schäfer, A., Horn, H. & Ahlrichs, R. (1992). *J. Chem. Phys.* **97**, 2571–2577.
Steiner, T. (2000). *Acta Cryst.* B**56**, 673–676.
Tripos (2007). *SYBYL*, Version 8.0. Tripos Inc., St Louis, Missouri, USA, http://www.tripos.com.

Torii, K. & Iitaka, Y. (1973). *Acta Cryst.* B**29**, 2799–2807.
Tulip, P. R. & Clark, S. J. (2005). *Phys. Rev. B*, **71**, 195117.
Ugliengo, P., Viterbo, D. & Chiari, G. (1993). *Z. Kristallogr.* **207**, 9.
Weissbuch, I., Frolow, F., Addadi, L., Lahav, M. & Leiserowitz, L. (1990). *J. Am. Chem. Soc.* **112**, 7718–7724.
Wu, Y.-F., Li, F.-F. & Jin, L.-F. (2005). *Acta Cryst.* E**61**, o3752–o3753.
Wu, Y.-F., Xiao, F.-P., Jin, L.-F., Li, F.-F. & Dai, X.-Y. (2005). *Acta Cryst.* E**61**, o4028–o4029.
Yamashita, M., Inomata, S., Ishikawa, K., Kashiwagi, T., Matsuo, H., Sawamura, S. & Kato, M. (2007). *Acta Cryst.* E**63**, o2762-o2764.