

Database guided conformation selection in crystal structure prediction of alanine

Timothy G. Cooper,^a William Jones,^a W. D. Samuel Motherwell^b and Graeme M. Day^{*a}

Received 12th February 2007, Accepted 22nd March 2007

First published as an Advance Article on the web 4th April 2007

DOI: 10.1039/b702136d

Crystal structure prediction calculations have been performed for the α -amino acid alanine with the intention of developing reliable computational methods for flexible molecules and, specifically, to study the crystal packing of the more flexible amino acids. For the α -amino acids, the density functional theory geometry optimised conformations of the isolated molecules are considerably different, in both geometry and form, to what is observed in the crystal structures. The molecules take the zwitterionic form in the observed crystals, but are nonionised for the isolated molecules. The quantum mechanically optimised structure of the isolated molecule is therefore a poor starting point for computationally generating putative crystal structures. We show that, by limiting the conformations of alanine to the torsion angle distributions in the observed crystal structures of similar molecules in the Cambridge Structural Database, sets of likely crystal structures can be generated, with the lowest energy racemic and enantiopure crystal structures corresponding to the experimentally observed crystal structures.

Introduction

Crystal structure prediction (CSP) is becoming an increasingly important tool in the understanding and anticipation of polymorphs of organic molecular crystals. With advances in the modeling of intermolecular interactions and in algorithms for sampling crystal packing phase space, the reliability of CSP methods for rigid molecules is continually improving. A recurring problem in CSP, however, is that for a given molecule, the energies of many distinct crystal structures are often very close, *i.e.* typical energy differences are less than 1 kJ mol⁻¹. The relative energies, which are used to judge the stability of computer-generated crystal structures, are sensitive to small changes in the model potential and the assumed geometry of the molecule, which is the building block used to computationally generate putative crystal structures.

For small rigid molecules, the crystalline environment does not significantly affect the molecular geometry, so that a quantum mechanically optimised molecular structure can be used throughout the calculations. Molecular flexibility introduces two major problems for CSP calculations: (i) the increased search space, where each degree of flexibility (*e.g.* rotation about a single bond) adds a dimension to the search space, and (ii) the evaluation of relative intramolecular energies as well as intermolecular energies between structures, which need to be both accurate and balanced. At the moment, crystal structure prediction for conformationally flexible molecules is unreliable, as is clearly highlighted in the blind tests of crystal structure prediction;¹ even small variations in

molecular geometry, *e.g.* amide pyramidalisation,² can greatly affect the prediction of observed crystal structures.

With these limitations in mind, we are exploring approaches to extend the applicability of current CSP methods to molecules of increasing flexibility. Reliable computational methods for predicting the crystal structures of flexible molecules would be particularly important in making an impact on pharmaceutical solid form development, where molecules of interest are much larger than those whose crystal structures can currently be reliably predicted. One approach for considering molecules with relatively little flexibility (one or two flexible torsional angles) is to consider a set of molecular geometries, chosen from an analysis of the isolated molecule.^{3–5} These molecular conformations, which might be chosen from local minima on the conformational energy surface, are used to generate sets of crystal structures that are then lattice energy minimised, either keeping the molecular conformation fixed^{3,5} or allowing it to relax under the influence of crystal packing forces.^{4,6} The main problem with molecules containing more than a couple of flexible torsional angles is the large conformational space that must be considered. Another approach for choosing the molecular conformations of flexible molecules might involve using information from crystal structures of similar molecules, guiding the modeling work towards those conformations that are most likely to be observed. We are investigating how information in the Cambridge Structural Database⁷ can be combined with modeling methods to predict crystal structures of otherwise difficult systems.

To develop methods for predicting crystal structures of flexible molecules, we have chosen to study the crystal packing of the α -amino acids. The α -amino acids represent a series of flexible molecules whose crystal structures are reasonably well studied, giving a testing ground for the development of

^aThe Pfizer Institute for Pharmaceutical Materials Science, Department of Chemistry, University of Cambridge, Cambridge, UK.
E-mail: gmd27@cam.ac.uk; Fax: +44 (0)1223 336362;
Tel: +44 (0)1223 336390

^bCambridge Crystallographic Data Centre, Cambridge

methods of handling larger, more flexible molecules in CSP. Furthermore, members of large families of chemically similar molecules are particularly amenable to the use of information in the Cambridge Structural Database (CSD), since there are likely to be a number of known crystal structures of related molecules that will provide useful information on preferred conformations.

In addition to problems associated with conformational flexibility, the amino acids provide an extra challenge in that they are known to crystallise in the zwitterionic form, even though the nonionised form is the most stable in the gas phase.^{8,9} As a result, gas phase optimisation will not provide a suitable molecular structure for crystal structure prediction, and additional constraints are needed to maintain the zwitterionic form. Furthermore, for the α -amino acids, due to the proximity of the carbonyl and amino groups, there is likely to be competition between intra- and intermolecular hydrogen bonding in the crystal structures. A balance between the two will then dictate the structure of the observed crystals.

CSP of glycine, the simplest α -amino acid, has been reported by our group,¹⁰ with the known polymorphs being reasonably well predicted using a rigid molecular geometry obtained from density functional theory. Alanine is the simplest chiral amino acid, with a methyl group attached to the α -carbon atom, and both the enantiopure and racemic forms have crystal structures in the CSD. They are each observed to crystallise in only one form and the observed structures are surprisingly similar, with nearly identical cell dimensions and conformations of the alanine molecules. L-Alanine crystallises in the $P2_12_12_1$ space group with pairs of layers anti-parallel to one another (Fig. 1a). Each layer is significantly stabilised by characteristic head-to-tail hydrogen bonds generated by a cell translation and a 2_1 screw axis. DL-Alanine crystallises in the $Pna2_1$ space group. This is a rare case of an amino acid racemate crystallising in a non-centrosymmetric space group (DL-tyrosine also crystallises in $Pna2_1$), although the $Pna2_1$ space group is one of the most commonly observed space groups and, as such, is generally included in crystal structure predictions. The observed structure of DL-alanine consists of alternating layers each containing only one type of isomer (Fig. 1b), and, as in the case of L-alanine, layers are stabilised by head-to-tail hydrogen bonds. Adjacent layers are interconnected by head-to-tail sequences generated by a glide plane. Both known crystal structures were located amongst the lowest energy structures in a previous CSP study¹¹ using a semirigid molecular model; the dihedral angles of alanine were allowed to vary, with the bond lengths and angles kept fixed. However, the search was limited to the two known space groups, with the other common space groups not considered. Despite their considerable biological importance and because of the relative flexibility of the larger amino acids, there have been, to our knowledge, no other attempts at predicting the crystal structures of the α -amino acids. Here, we are developing and evaluating methods using the prediction of the observed crystal structures of alanine as a target, considering all the most commonly observed space groups. The aim is to develop methods that can be extended to the more complex amino acids and systems whose crystal structures are unknown.

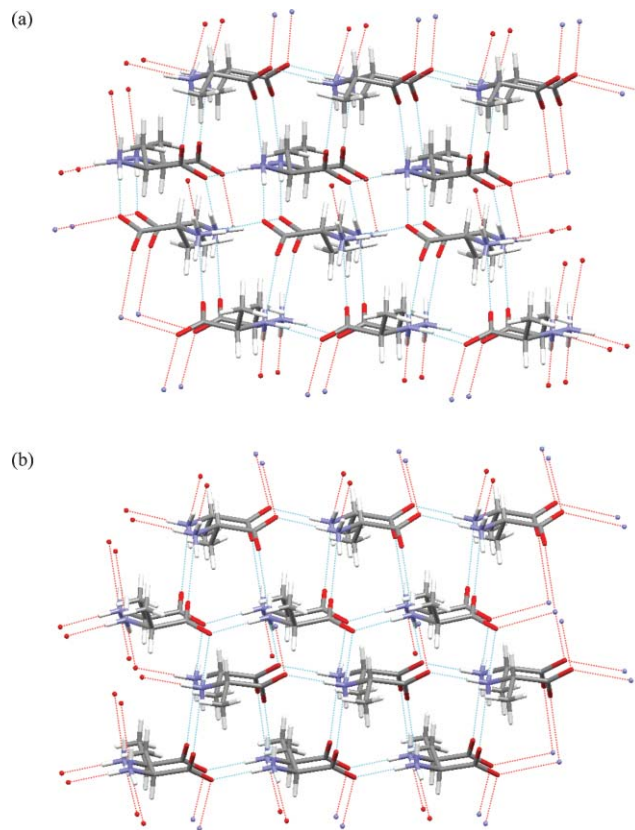


Fig. 1 Packing diagrams of the observed structures of (a) L-alanine showing the characteristic head-to-tail hydrogen bonding with pairs of layers anti-parallel to one another, and (b) DL-alanine with layers containing only one type of isomer and all layers parallel. Hydrogen bonds are shown by dotted lines.

Methods

Cambridge structural database searches

A survey of the torsion angles of all the available $Z' = 1$ chiral α -amino acids in the Cambridge Structural Database (CSD) was performed using ConQuest 1.8.¹² The carboxylate and amino torsional angles were measured in the best representation of each available crystal structure (either structures obtained from neutron diffraction or the structure with the lowest R factor); both torsion angles are measured relative to the α -hydrogen atom ($\tau_1 = \text{HNCH}$, $\tau_2 = \text{OCCH}$, Fig. 2).

Molecular structure calculations

For geometry optimisations and torsion angle scans of the molecule in the gas phase we performed density functional

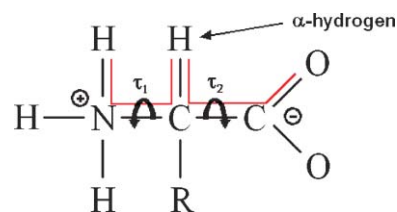


Fig. 2 Structure and torsional angle definition of the zwitterionic form of a general α -amino acid.

theory (VWN/DNP) calculations using the Dmol¹³ module within the Accelrys Material Studio package. Full geometry optimisation leads to the non-zwitterionic form, so it was necessary to constrain the molecule to be zwitterionic by fixing N–H bondlengths at 1.035 Å. Apart from τ_1 and τ_2 , which were constrained at a series of values in the ranges $120^\circ \leq \tau_1 \leq 240^\circ$, $0^\circ \leq \tau_2 \leq 180^\circ$ and incremented in 10° steps, all other degrees of freedom were allowed to relax during the torsion angle scans.

Crystal structure searches

Crystal structure searches were performed with the CrystalPredictor code, which uses a low-discrepancy sequence (a near-uniformly distributed quasi-random sequence) to generate candidate crystal structures with quasi-random values for unit cell dimensions, molecular orientations and positions¹⁴ followed by rigid molecule lattice energy minimisation. Searches were performed with a set of rigid molecular geometries, chosen to sample the relevant region of the conformational energy surface. Preliminary searches were performed in three space groups, for which a total of 20000 crystal structures were energy minimised for each molecular conformation, while the final searches of twelve space groups were continued until 50 000 energy minimisations had been performed for each conformation. We monitored convergence of the set of low energy structures and the 20 000 and 50 000 minimisations were sufficient to give a complete sampling in all cases.

The initial searches using the DFT geometry optimised and the observed conformations of alanine were run in the two observed space groups ($P2_12_12_1$ and $Pna2_1$) and $P2_1/c$. $P2_12_12_1$ and $Pna2_1$ to see whether the observed crystal structures were generated and how their energies compared to the other possibilities, and $P2_1/c$, the most commonly observed space group for molecular crystals, to get a wider view of possible crystal structures and their energies. The final crystal structure searches were extended to 12 of the most common space groups: 5 chiral space groups ($P2_12_12_1$, $P2_1$, $P1$, $P2_12_12$ and $C2$) and 7 racemic space groups ($P2_1/c$, $Pna2_1$, $Pnma$, $C2/c$, $P\bar{1}$, $Pbca$ and $Pbcn$), which account for over 90% of all the entries found in chiral and racemic space groups in the CSD.¹⁵ Since there is little variation in the value of τ_1 amongst the α -amino acid crystal structures, we fixed it at a value of 180° , which is the mean value of the distribution of observed conformations, and scanned through the full range of observed τ_2 values. We considered values of τ_2 every 10° in the range $10^\circ \leq \tau_2 \leq 80^\circ$ and then combined the predictions from each molecular conformation. Structures were then clustered based on structure to remove duplicates.

Clustering of structures

To compare the predicted and observed crystal structures and for removal of duplicate structures (“clustering”), we used the Compact¹⁶ algorithm with a tolerance of 20% on interatomic distances (excluding hydrogen atoms) and a cluster of 15 molecules (*i.e.* one molecule and a coordination sphere of its 14 nearest neighbours).

Some crystal packings are found in searches using more than one molecular model (*i.e.* different values of τ_2), with small

differences in packing caused by differences in molecular geometry. The 20% tolerance on interatomic distances is sufficiently relaxed to cluster such structures (*i.e.* treating them as identical). The structure with the lowest total (lattice + conformational) energy is kept, and, in this way, the best molecular conformation is retained for each computer-generated crystal structure.

Model potentials

For the lattice energy minimisation of the structures from CrystalPredictor, we used an exp-6 model potential with C, N, O and H_c (hydrogen bonded to carbon) parameters parameterised by Williams and coworkers.¹⁷ Parameters to describe H_n (hydrogen bonded to nitrogen) were taken from ref. 18. During the generation of crystal structures, electrostatic interactions were modelled by electrostatic potential derived partial charges obtained from a single point DFT calculation of the gas phase alanine molecule (VWN/DNP). For the final energy minimisations, we applied the same exp-6 model potential with the electrostatic interactions described using a more elaborate distributed multipole model. Atomic multipoles, up to hexadecapole on each atom, were taken from a distributed multipole analysis¹⁹ (DMA) of the electron density (B3LYP/6-31G**), calculated for the specific molecular conformation under consideration using the CADPAC electronic structure program.²⁰ Final minimisations were performed using the DMAREL²¹ crystal structure modelling program; all exp-6 interactions were summed to a 15 Å cutoff, charge-charge, charge-dipole and dipole-dipole electrostatic interactions were summed using the Ewald summation, with all higher order electrostatic interactions (up to R⁻⁵) summed to a 15 Å cutoff on whole molecules.

Results and discussion

Preliminary searches using fully optimised and observed molecular conformations

Alanine has two flexible torsion angles that are important for its crystal packing: rotation of the amino group (τ_1) and the carboxylate group (τ_2), both of which we measure with respect to the α -hydrogen atom (Fig. 2). Unconstrained DFT geometry optimisation of the isolated alanine molecule led to proton transfer from the amino to the carboxylate group. Therefore, N–H bondlengths were fixed (at typical lengths of 1.035 Å) to keep the molecule in its zwitterionic form, resulting in an optimised structure (Fig. 3a) that was significantly different from the observed alanine conformation, which is nearly identical in the L- and DL-alanine crystal structures (Fig. 3b): $[\tau_1, \tau_2]_{\text{opt}} = [120^\circ, 61^\circ]$; $[\tau_1, \tau_2]_{\text{L-alanine}} = [176^\circ, 45^\circ]$ (measured in CSD refcode LALNIN12); $[\tau_1, \tau_2]_{\text{DL-alanine}} = [178^\circ, 45^\circ]$ (measured in DLALNI01). The difference is mainly due to the isolated molecule forming a strong intramolecular hydrogen bond, while the amine group twists in the crystal structures in order for all three of the amino hydrogen atoms to hydrogen bond to oxygen atoms of surrounding molecules, breaking the intramolecular hydrogen bond. The DFT isolated molecule calculations suggest that the energy for the observed conformation of alanine is 40 kJ mol⁻¹ less

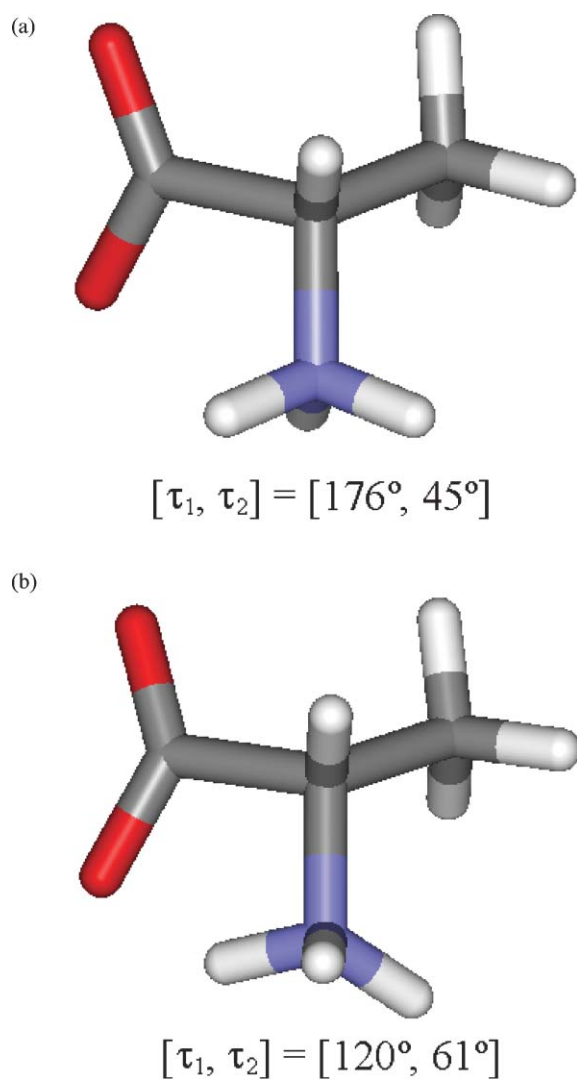


Fig. 3 Structure of (a) the observed conformation of L-alanine (LALNIN12) and (b) the DFT geometry optimised conformation.

stable than the DFT geometry optimised structure. Normally the observed conformation in molecular crystals is expected to be within a few kJ mol^{-1} of the geometry optimised isolated molecule²² – the energy difference here shows that alanine is a case where packing induces extreme changes in molecular conformation.

Predictions using the geometry optimised molecular conformation generate crystal structures matching both the observed structures of L-alanine (LALNIN12) and DL-alanine (DLALNI01) within the tolerances used by Compack. However, the differences in conformation of alanine in the crystal structures (*e.g.* Fig. 4) result in the predicted structures being poorly ranked in terms of lattice energy. Five enantiopure crystal structures up to 3.9 kJ mol^{-1} lower in energy than LALNIN12 were found in $P2_12_12_1$, and 27 racemic crystal structures lower in energy than DLALNI01 were found (5 in $Pna2_1$ and 22 in $P2_1/c$), with the lowest energy calculated to be 4.1 kJ mol^{-1} more stable than the observed crystal structure. These are unacceptable results, especially considering the limited space groups sampled, demonstrating

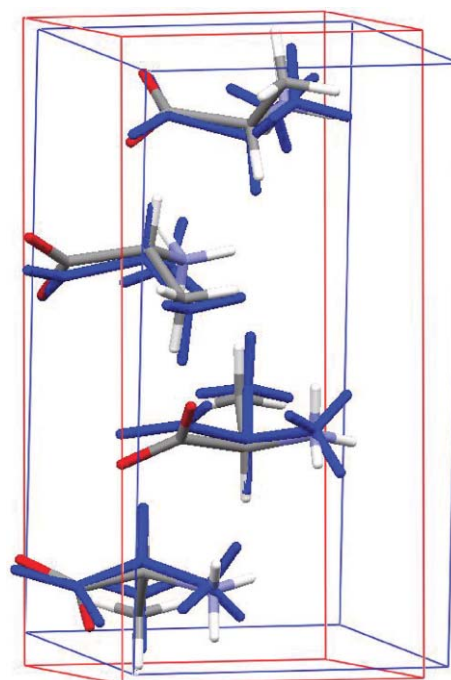


Fig. 4 Comparison of the observed (DLALNI01) and the corresponding predicted (blue) crystal structure using the DFT geometry optimised molecular conformation of alanine.

that the optimised geometry of the isolated alanine molecule is a poor choice for crystal structure prediction.

To test that if we were able to more accurately estimate the observed molecular conformation, we could better predict the observed structures, CSP calculations were performed using τ_1 and τ_2 fixed at values from the experimentally determined structure (CSD refcode LALNIN12). In this search, the observed crystal structures of both L and DL-alanine were the second most stable predicted structure for each system, with ΔE between the observed and lowest energy predicted structures of 1.2 and 1.3 kJ mol^{-1} , respectively. From these results, we concluded that the choice of molecular conformation is the major problem in the successful prediction of the known crystal structures of alanine and a closer approximation to the observed conformation is needed.

Distribution of conformations in the CSD and the calculated conformational energy surface

Distributions of τ_1 and τ_2 for the chiral α -amino acids in the CSD† (ignoring τ_1 for proline since it is part of a ring) revealed that there is relatively little variation in τ_1 across the amino

† Cambridge structural database REFCODES of the available α -amino acids: DLALNI01 (DL-alanine), LALNIN12 (L-alanine), LSERIN01 (L-serine (I)), LSERIN16 (L-serine (II)), DLSERN11 (DL-serine), VALIDL (DL-valine (I)), VALIDL02 (DL-valine (II)), LTHREO02 (L-threonine), LASPRT (L-aspartic acid), DLASPA02 (DL-aspartic acid), LCYSTN12 (L-cysteine), BOQCUF (DL-cysteine), LHISTD04 (L-histidine (I)), LHISTD13 (L-histidine (II)), DLHIST (DL-histidine), TACQUJ (DL-glutamine), GLUTAM01 (L-glutamine), LGLUAC03 (L-glutamic acid (α)), LGLUAC11 (L-glutamic acid (β)), YUYMOU (DL-glutamic acid), LTYROS11 (L-tyrosine), DLTYRS (DL-tyrosine), DLLEUC (DL-leucine), DLILEU02 (DL-isoleucine), QQQBTP02 (DL-tryptophan)

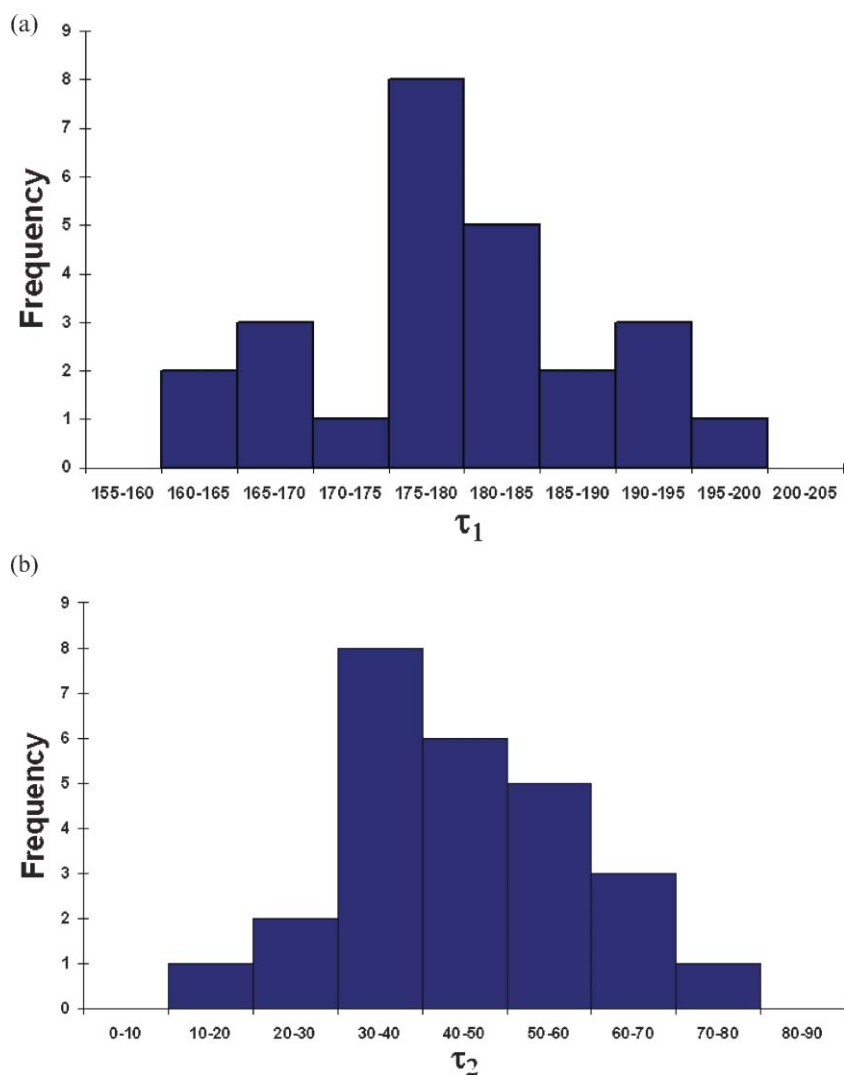


Fig. 5 Distributions of τ_1 and τ_2 for the $Z' = 1$ chiral α -amino acids in the CSD.

acids, with the torsion angles of all the amino acids falling within 20° of being fully staggered with respect to the α -hydrogen atom ($\tau_1 = 180^\circ$) (Fig. 5a). The distribution of τ_2 is more varied, but all the angles fall within the range 10 – 80° and over 70% between 30 and 60° (Fig. 5b). These distributions of τ_1 and τ_2 provide information on the conformational preferences of this family of molecules in the crystalline state, so we focused our attention on this region of conformational space in our alanine calculations.

The calculated conformational energy surface (Fig. 6) shows the region of molecular conformation for the α -amino acids in the CSD. A large portion of this region, and the region where most of the observed conformations are found, has conformational energies between 20 and 30 kJ mol^{-1} above the global minimum. However, the α -amino acids are generally crystallised from water, where the conformations will be close to those found in the solid state, so these are not a true indication of the energies that must be overcome upon packing for the molecules to adopt the observed conformations. Nevertheless, typically observed molecular conformations are within a few kJ mol^{-1} of the geometry optimised isolated

molecule, so the conformations adopted by all the α -amino acids in the observed crystal structures are exceptions to the rule. This further highlights that we cannot use conformations based solely on isolated molecule calculations of molecular geometry when attempting to predict their crystal structures.

Initial investigations concentrated on conformations of alanine around the mean values of τ_1 and τ_2 for the amino acids in the CSD (mean $\tau_1 = 179^\circ$ and $\tau_2 = 44^\circ$), generating crystal structures in the three space groups $P2_12_12_1$, $Pna2_1$ and $P2_1/c$ with alanine in fixed conformations. The values of the two torsional angles were then varied around these conformations to investigate how the ranking of structures generated by CSP depends on the molecular conformation, by comparing these computer-generated crystal structures with the two known alanine crystal structures. The relative energies and lattice energy rankings of DLALNI01 and LALNIN12 amongst the predicted racemic and enantiopure crystal structures are summarised in Table 1. The two observed structures are expected to be the lowest energy possibilities. The results show that, within the range of torsion angles found in the CSD, there is little dependence of the ranking of the

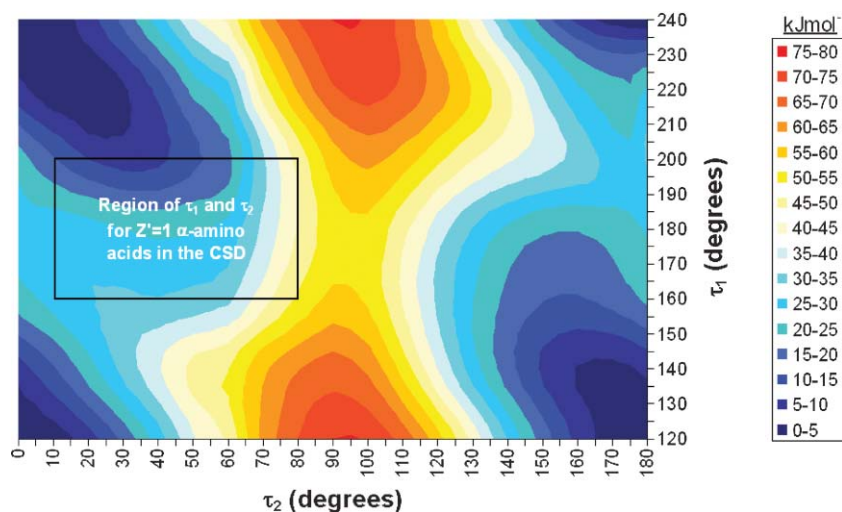


Fig. 6 Calculated conformational energy contour plot for L-alanine as a function of τ_1 and τ_2 .

observed crystal structures on τ_1 . For values of τ_2 close to the observed value ($\tau_2 = 45^\circ$), the observed crystal structures are ranked amongst the most stable predicted structures for a wide range of τ_1 (Table 1). In contrast, even for values of τ_1 very close to the observed value, large deviations in τ_2 lead to poor rankings of the observed crystal structures; this is especially

Table 1 Ranking and relative energies of the observed crystal structures from crystal structure searches with a range of rigid molecular geometries in the observed spacegroups for L-alanine ($P2_12_12_1$) and the observed spacegroup ($Pna2_1$) and $P2_1/c$ for DL-alanine

Torsion angles		DL-Alanine		L-Alanine	
τ_1	τ_2	N_{lower}^a	$\Delta E/\text{kJ mol}^{-1}$	N_{lower}^a	$\Delta E/\text{kJ mol}^{-1}$
175	10	not found ^b	—	not found ^b	—
180	10	6	5.3	0	-3.4
185	10	not found ^b	—	not found ^b	—
170	20	5	4.5	0	-1.0
175	20	6	5.2	0	-3.0
180	20	8	6.8	0	-3.3
185	20	11	7.1	0	-1.1
150	40	0	-4.0	0	-6.5
155	40	0	-2.5	0	-7.0
160	40	0	-4.7	0	-7.9
165	40	0	-5.5	0	-7.8
170	40	0	-5.8	0	-8.2
175	40	0	-5.1	0	-5.9
180	40	0	-4.6	0	-4.9
185	40	0	-2.5	0	-5.0
190	40	1	0.1	0	-3.9
195	40	2	1.0	0	-3.7
200	40	not found ^b	—	not found ^b	—
170	60	0	-1.1	1	1.1
175	60	0	-1.3	1	1.9
180	60	1	0.3	1	3.3
185	60	6	2.3	2	2.6
175	70	20	6.3	3	9.1
180	70	26	8.1	5	10.9
185	70	20	7.2	3	10.3

^a N_{lower} is a count of the number of predicted crystal structures with lower calculated energies than the observed form. ^b No predicted structure corresponds to the observed crystal structure within the tolerances on interatomic distances used in the COMPACT comparison algorithm.

evident for DL-alanine, where many predicted crystal structures become more stable than the observed structure at extreme values of τ_2 (Table 1).

CSP in all the most common spacegroups

The results of these preliminary calculations were used to choose a set of conformations with which to generate crystal structures in all of the common space groups for molecular organic crystal structures. The calculations demonstrated little dependence of the predictions on τ_1 , so we chose to fix its value at 180° , which is the mean value of τ_1 for the chiral α -amino acids in the CSD. However, it seems necessary to allow variation in τ_2 , so we generated structures with 8 separate values of τ_2 , at 10° intervals in the range found in the CSD ($10^\circ \leq \tau_2 \leq 80^\circ$).

The observed structure of DL-alanine was located in all 8 searches, within the tolerances on interatomic distances used in the COMPACT comparison. The observed structure of L-alanine was found in all but the search with conformation $[180^\circ, 10^\circ]$, although a match to the observed structure is found when the interatomic distance tolerance is increased to 30%. The rankings within the set of predictions for each conformation are summarised in Table 2. As with the preliminary search, when τ_2 is close to the observed torsion angle (45°) both observed crystal structures are found to be the most stable predicted structures.

Many of the predicted crystal structures are found in several of the searches, with small differences owing to the variation in τ_2 . To generate our final set of putative crystal structures, the lists generated in each individual search were combined and the entire set was clustered using a 20% tolerance on interatomic distances, which removes structures with the same crystal packing, with the structure of each cluster having the lowest total (conformational + lattice) energy retained. Conformational energies were evaluated from the DFT calculations and combined with the exp-6 + DMA calculated lattice energies. This resulted in a final set of over 2000 distinct crystal structures, including both enantiopure and racemic structures, within 20 kJ mol^{-1} of the global minimum total

Table 2 Ranking and relative energies of the observed crystal structures for L- and DL-alanine. Searches were completed in 5 chiral and 7 racemic spacegroups with τ_2 scanned in the range $10^\circ \leq \tau_2 \leq 80^\circ$ and τ_1 kept fixed at 180°

τ_2	DL-Alanine		L-Alanine	
	ΔE	N_{lower}^a	ΔE	N_{lower}^a
10	5.3	14	-3.4	0
20	6.8	17	-3.3	0
30	0.2	1	-6.0	0
40	-4.6	0	-4.9	0
50	-4.9	0	-1.6	0
60	0.3	2	3.3	2
70	10.0	66	10.9	9
80	15.2	175	not found ^b	—

^a N_{lower} is a count of the number of predicted crystal structures with lower calculated energies than the observed form. ^b No predicted structure corresponds to the observed crystal structure within the tolerances on interatomic distances used in the CompPack comparison algorithm.

energy, each different crystal packing with its optimum molecular conformation. Fig. 7a and 7b summarise the results for the pure enantiomeric and racemic crystal structures respectively.

LALNIN12 and DLALNI01 were located in the final set of structures and found to be the most stable predicted structures

for both the pure enantiomeric and racemic systems, 2.0 and 1.7 kJ mol⁻¹ more stable than the next lowest energy structures respectively. The structures of the lowest energy predictions and observed crystal structures are compared in Table 3; the observed cell parameters are predicted to within 2.5%. Both observed crystal structures have lowest total energy with $\tau_2 = 40^\circ$. Of note, had we selected to scan τ_2 with coarser intervals of 20° , e.g. $\tau_2 = (10^\circ, 30^\circ, 50^\circ, 70^\circ)$, the observed crystal structures of L-alanine and DL-alanine would still have been found as the most stable structures within the final total-energy ranked lists, with $\tau_2 = 50^\circ$. Such coarser sampling of torsion angles might be necessary for molecules with more degrees of freedom.

Conclusions

We found that alanine is an especially challenging molecule for CSP, largely due to the striking changes in molecular structure upon crystallisation. Consequently, unlike many small molecules, the quantum mechanically optimised structure of the isolated molecule is not a good choice for generating the likely crystal structures. The predicted versions of the known alanine crystal structures using the optimised molecular structure are distorted and poorly ranked in energy relative to the other packing possibilities.

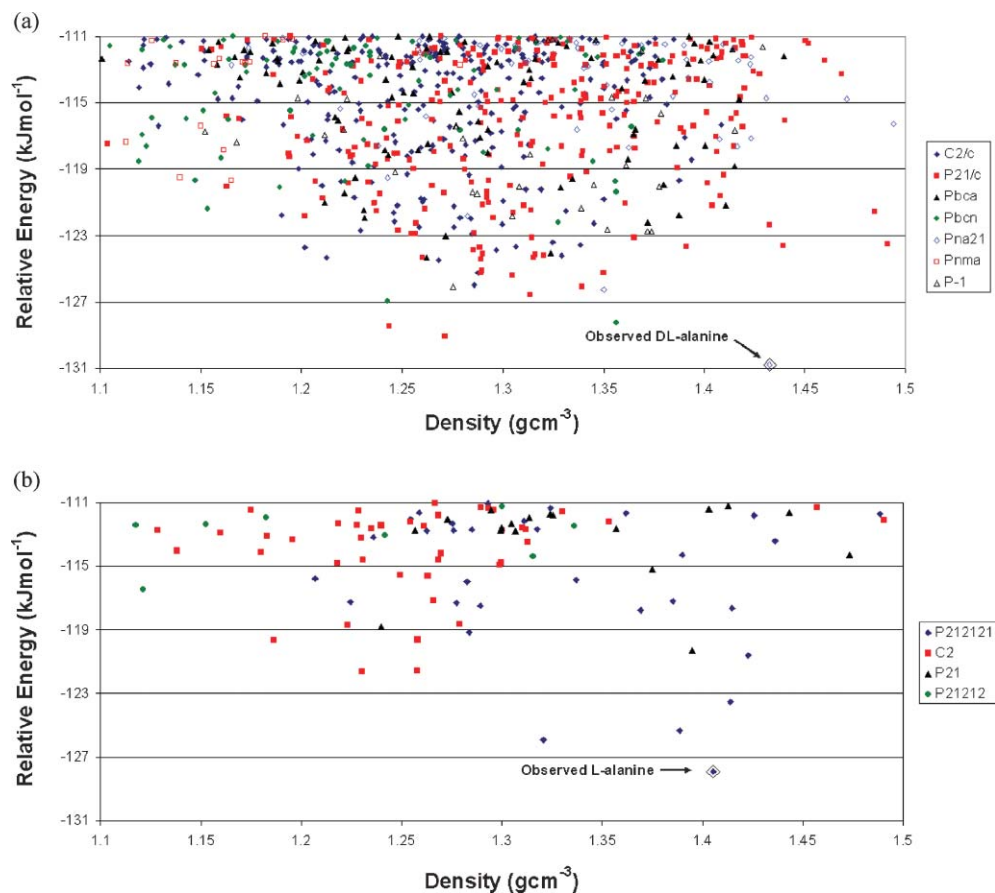


Fig. 7 Total energy (lattice + relative conformational energy) vs density plots of the combined crystal structure predictions fixing τ_1 at 180° and varying $10^\circ \leq \tau_2 \leq 80^\circ$ for (a) the 7 most common racemic space groups, and (b) the 5 most common chiral space groups (no low energy structures were found in space group *P1*). Each point is a distinct crystal structure and the observed crystal structures of DL-alanine and L-alanine are highlighted.

Table 3 Structural parameters for the observed crystal structures and the best predicted crystal structures of alanine, with percentage errors in brackets

Structure	<i>a</i> /Å	<i>b</i> /Å	<i>c</i> /Å	(Volume/molecule)/Å ³	Density/g cm ⁻³	τ ₁ /°	τ ₂ /°
DLALNI01	12.026	6.032	5.829	422.86	1.399	178	45
Predicted DL-alanine	12.218 (+1.6%)	5.933 (-1.6%)	5.699 (-2.2%)	413.09 (-2.3%)	1.433 (+2.4%)	180	40
LALNIN12	6.025	12.324	5.783	429.40	1.378	176	45
Predicted L-alanine	6.012 (-0.2%)	12.392 (+0.6%)	5.652 (-2.3%)	421.10 (-1.9%)	1.405 (+2.0%)	180	40

In cases like this, where calculations on the isolated molecule cannot provide an adequate molecular conformation for CSP, the conformational preferences of the molecule in the crystal can be predicted by examining the crystal structures of similar molecules in the Cambridge Structural Database. Here, a survey of the α -amino acids in the Cambridge Structural Database proved very useful in reducing the conformational space that needs to be considered. There is a narrow range of τ_1 in these crystal structures centered around 180° and a wider distribution of τ_2 , ranging from 10 to 80°. These conformational preferences in the crystal structures of the α -amino acids helped direct our choice of molecular models used in the CSP calculations; the search for crystal structures in this limited conformational space located the two known crystal structures of alanine (*Pna*₂₁ DL-alanine and *P2*₁*2*₁*2*₁ L-alanine) as the lowest energy racemic and enantiopure possibilities, respectively. Furthermore, no other crystal structures are found within about 2 kJ mol⁻¹ of either known crystal structure.

The results presented here for alanine demonstrate how useful the Cambridge Structural Database can be as a tool for reducing the conformational space that needs to be considered for a molecule in crystal structure prediction calculations. The combination of modeling and database methods should be particularly useful for molecules where the crystalline environment is crucially important in determining the molecular conformation. This approach might also be necessary as molecules of interest become more complex, resulting in larger conformational space and pushing the limits of molecular size for accurate quantum mechanical studies. We are now extending our studies to the larger amino acids, many of which crystallise with two independent molecules in the asymmetric unit (*Z'* = 2), so the reduction of conformational space to be considered is vital for computational studies of these systems. The choice of structures from the CSD that are used for guiding the conformations is clearly important; the set of structures must be large enough to be meaningful, but specific enough to be relevant to the system of interest. Further studies are required to investigate how to choose the best set of structures from the CSD for a general flexible molecule.

Acknowledgements

We thank Dr Panos Karamertzanis for the installation, use and maintenance of the CrystalPredictor code, Dr

Neil Feeder for valuable discussions and the Pfizer Institute for Pharmaceutical Materials Science for funding.

References

- 1 G. M. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, B. P. van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, A. Torrisi, H. Scheraga, Y. A. Arnautova, M. U. Schmidt and P. Verwer, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2005, **61**, 511.
- 2 A. J. Cruz Cabeza, G. M. Day, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2006, **6**, 1858.
- 3 C. Ouvrard and S. L. Price, *Cryst. Growth Des.*, 2004, **4**, 1119.
- 4 G. M. Day, W. D. S. Motherwell and W. Jones, *Phys. Chem. Chem. Phys.*, 2007, **14**, 1693.
- 5 H. Nowell and S. L. Price, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2005, **61**, 558.
- 6 P. G. Karamertzanis and S. L. Price, *J. Chem. Theory Comput.*, 2006, **2**, 1184.
- 7 F. H. Allen, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 380.
- 8 J. Almlöf, Å. Kvik and J. O. Thomas, *J. Chem. Phys.*, 1973, **59**, 3901.
- 9 K. Iijima, K. Tanaka and S. Onuma, *J. Mol. Struct.*, 1991, **246**, 257.
- 10 G. M. Day, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2005, **5**, 1023.
- 11 V. E. Bazterra, B. B. Ferraro and J. C. Facelli, *Int. J. Quantum Chem.*, 2003, **96**, 312.
- 12 I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson and R. Taylor, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 389.
- 13 B. Delley, *J. Chem. Phys.*, 1990, **92**, 508.
- 14 P. G. Karamertzanis and C. C. Pantelides, *J. Comput. Chem.*, 2005, **26**, 304.
- 15 <http://www.ccdc.cam.ac.uk/products/csd/statistics/>.
- 16 J. Chisholm and W. D. S. Motherwell, *J. Appl. Crystallogr.*, 2005, **38**, 228.
- 17 (a) S. R. Cox, L.-Y. Hsu and D. E. Williams, *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Cryst.*, 1981, **37**, 293; (b) D. E. Williams and S. R. Cox, *Acta Crystallogr., Sect. B: Struct. Sci.*, 1984, **40**, 404.
- 18 D. S. Coombes, S. L. Price, D. J. Willock and M. Leslie, *J. Phys. Chem.*, 1996, **100**, 7352.
- 19 A. J. Stone and M. Alderton, *Mol. Phys.*, 1985, **56**, 1047.
- 20 R. D. Amos, CADPAC, version 6.0, 1995; (with contributions from I. L. Alberts, J. S. Andrews, S. M. Colwell, N. C. Handy, D. Jayatilaka, P. J. Knowles, R. Kobayashi, N. Koga, K. Laidig, P. E. Maslen, C. W. Murray, J. E. Rice, J. Sanz, E. D. Simandiras, A. J. Stone and M. D. Su).
- 21 D. J. Willock, S. L. Price, M. Leslie and C. R. A. Catlow, *J. Comput. Chem.*, 1995, **16**(5), 628.
- 22 F. H. Allen, S. E. Harris and R. Taylor, *J. Comput.-Aided Mol. Des.*, 1996, **10**, 247.