ARTICLE

Amino acid hydrogen oxalate quasiracemates – sulfur containing side chains

moReceived 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Russell G. Wells,[a] Katriel D. Sahlstrom,[a] and Kraig A. Wheeler[a*]

New additions to quasiracemic materials have been developed by cocrystallizing a ternary component – hydrogen oxalate – with pairs of amino acid quasienantiomers where at least one of the side-chain *R* groups contains a sulfur atom. Of the eight quasiracemates investigated, six exhibit crystal packing that drastically deviates from the expected centrosymmetric alignment present in the racemic counterparts and the extant database of quasiracemic materials. These structures were quantitatively assessed for conformational similarity (*CCDC-Mercury* structure overlay) and the degree of inversion symmetry (*Continuous Symmetry Measures*) for each quasienantiomeric pair. Despite the variance in quasienantiomeric components, these structures exhibit a high degree of isostructurality where the principal components assemble by a complex blend of common N^{+}-H···O and O-H···O^{-} interactions. These charge-assisted hydrogen-bonded networks form thermodynamically favored crystal packing that promotes cocrystallization of a structurally diverse set of quasienantiomeric components.

Introduction

The companion article preceding this report highlights the utility of oxalic acid as a designer additive for organizing pairs of amino acid quasienantiomeric components.[1] In that investigation, we showed how amino acids decorated with aliphatic side chains (*R*) when combined with hydrogen oxalate (*i.e.,* Nva·Ile·2HOx, Nva·Leu·2HOx, Leu·Ile·2HOx, and Leu·Nle·2HOx) provided significant opportunities to understand the structural effects coformer molecules play on the self-assembly process of quasiracemates. Outcomes from that report contribute to a growing body of work that explores the cocrystallization tendencies from pairing chemically distinct L- and D-amino acid components. When considering the 32 known examples of coformer-free amino acid quasiracemates, these collective crystal structures offer an important view of the structural variation possible when combining equimolar ratios of amino acids.[2-10] While it might be anticipated that component pairs with similar topological features such as L-methionine·D-norleucine[3] and L-isoleucine·D-leucine[5] form quasiracemic crystalline phases, the success of using quasienantiomers of greater structural diversity (*e.g.,* L-isoleucine·D-alanine[5] and L-phenylalanine·D-norvaline[2]) was less predictive. By examining a new set of ternary amino acid quasiracemates, these studies underscored the diversity of amino acid quasienantiomers, the potential importance of ionic hydrogen bonds to crystal lattice stabilization, and they paved the way for further studies on amino acid quasiracemates.

Our previous article in this series utilized a set of structural tools to quantitatively assess the conformational similarity and degree of inversion symmetry of the hydrogen oxalate and related non-hydrogen oxalate amino acid quasiracemates. For these structures where the *R* groups contain only C/H chemical functions, the molecular topologies for the pair of amino acids (*i.e.,* L-X·D-X') closely correlate, as do their crystal packing patterns to true inversion symmetry. The shared structural feature of these hydrogen oxalate structures, including the previous family of 32 quasiracemates, is that they all form crystal patterns that closely approximate inversion symmetry.

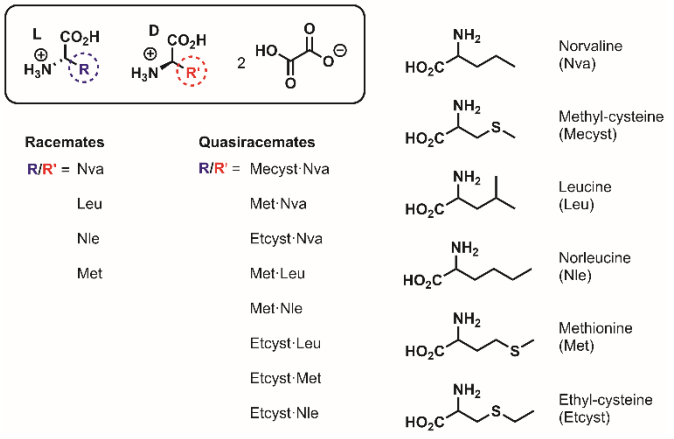

Fig. 1 Racemic and quasiracemic hydrogen oxalate amino acid systems examined in the current study.

[a] *Department of Chemistry, Whitworth University, 300 West Hawthorne Road, Spokane, Washington, 99251, USA. E-mail: kraigwheeler@whitworth.edu.*

† Electronic Supplementary Information (ESI) available: Crystal growth procedures, X-ray crystallographic assessments, hydrogen-bond tables, and crystal packing diagrams of the racemic compounds, structure indices (χ_{RMS} and C_i), CSD search data, and dihedral angle distributions. CCDC 2017245, 2017246, 2017248-2017250, 2017252, 2017257, 2017258, 2017260, 2051774-2051777, 2054514. For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/x0xx00000x

In this article, we continue our effort to understand the structural consequence of using oxalic acid as a coformer molecule with amino acid quasiracemates, where at least one of the amino acid components includes a sulfur-containing *R* group (*i.e.,* methionine (Met), ethyl cysteine (Etcyst), and methyl cysteine (Mecyst)). As shown in Fig. 1, of the six building blocks selected for this study - norvaline (Nva), methyl cysteine, leucine (Leu), norleucine (Nle), methionine, ethyl cysteine - five contain straight chain side groups. Combining pairs of the L and D forms of these components with oxalic acid produced eight new quasiracemic crystal structures (L-Mecyst·D-Nva·2HOx, L-Met·D-Nva·2HOx, L-Etcyst·D-Nva·2HOx, L-Met·D-Leu·2HOx, L-Met·D-Nle·2HOx, L-Etcyst·D-Leu·2HOx, L-Etcyst·D-Met·2HOx, L-Etcyst·D-Nle·2HOx). In general, chain length, branching pattern, and the -S-/-CH_2- substitution provided the primary structural points of difference for these systems. While the -S-/-CH_2- substitutions are known isomorphic groups that have guided previous quasiracemate studies[11,12], the structural outcomes from using these functional groups in this study were profoundly unexpected since many of these quasiracemates formed decidedly non-centrosymmetric crystal packing patterns.

Results and discussion

Crystallographic Assessments

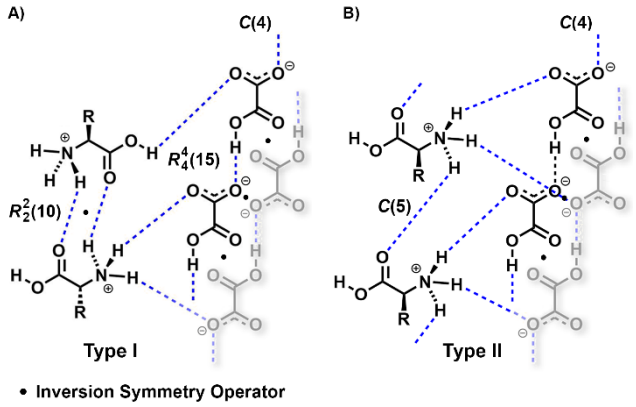

Fig 2. Illustrations of Type I and II hydrogen bond motifs for racemic and quasiracemic amino acid hydrogen oxalates. (Fig. 2A taken from ref. 1)

Racemates. Numerous homochiral[13-20] and racemic[21-28] crystal structures of amino acid hydrogen oxalates exist in the literature. A common theme from inspection of these structures is that the hydrogen oxalate components prefer to form robust hydrogen-bonded strands *via* O-H···O^- interactions. The amino acids then hydrogen bond to these hydrogen oxalate motifs using a complex blend of contacts. For this study, we initially focused our attention on the racemic forms of amino acid hydrogen oxalates where at least one of the components relates to the quasienantiomers included in this study. The structures of DL-Nva·HOx and DL-Met·HOx were prepared and crystallographically determined with DL-Leu·HOx[29] and DL-Nle·HOx[1] retrieved from the literature. Each component associated with these racemic systems crystallized in space group $P\bar{1}$ with Z' = 1 (Table S1, ✝). The structural similarity is further expressed in unit cell constants where the parameters a, b, c, α, β, and γ are noticeably similar for this set of structures. The hydrogen oxalate anions link to translationally related HOx neighbors *via* O-H···O^- interactions to give an extended motif with $C(5)$ graph set notation (Fig. 2).[30,31] These hydrogen oxalate chains self-assemble as inversion related dimers with close stacking between the chains (*i.e.,* 2.97 – 3.15 Å) promoted by a single amino acid NH_3^+ group *via* two N^+-H···O_{HOx} contacts. By evaluating the orientation of the amino acid···amino acid interactions, these structures could be further classified as Type I or II. The structures of DL-Nva·HOx, DL-Leu·HOx, and DL-Nle·HOx follow Type I patterns where pairs of amino acids assemble into $\boldsymbol{R_2^2}(10)$ centrosymmetric dimers aided by N-H···O=$C_{Amino\ Acid}$ contacts using the third NH_3^+ hydrogen atom (Fig. 2A). The remaining hydrogen-bond donor participates in $CO_2H_{amino\ acid}$···O^-_{HOx} contacts that control the spatial proximity between adjacent hydrogen oxalate dimer motifs with distances corresponding to the length of the crystallographic *b* axes (9.40 - 9.56 Å). This collection of hydrogen-bonded motifs creates a bilayer structure where the R group side chains and core amino acid residues reside in distinct *ab*-planes.

The methioninium hydrogen oxalate (DL-Met·HOx) structure exhibits Type II hydrogen-bond motifs. In addition to using the methioninium NH_3^+ group to span the hydrogen oxalate dimer strands *via* two N^+-H···O_{HOx} contacts, the third NH_3^+ hydrogen atom participates in N^+-H···O=$C_{methioninium}$ contacts linking $C(5)$ translationally related methioninium molecules (Fig. 2B). For Type II motifs and with the structure of DL-Met·HOx, the remaining methioninium carboxyl donor contributes to crystal packing *via* a $CO_2H_{methioninium}$···O^-_{HOx} interaction that effectively links the hydrogen oxalate dimer stacks with a distance of 9.44 Å. Type II motifs, similar to Type I, also form bilayer structures where the *R* groups are isolated in the *ab*-plane.

These four racemic HOx structures offer a valuable view of the structural tendencies of amino acid hydrogen oxalate systems. One such structural preference relates to the hydrogen oxalate anion as a key participant in supramolecular assembly. This small but structurally potent coformer molecule consistently self-assembles to generate pairs of centrosymmetrically related hydrogen-bonded strands. These HOx pillars form reliable spacing in the crystal at ~9.5 Å, where the inter-pillar space is occupied by the amino acid components. Given the structural irregularities described in the preceding sections for sulfur-containing amino acid hydrogen oxalate quasiracemates, it is important to note that the sulfur atom in DL-Met·HOx, as well as the C/H groups of the side chains in the other racemic structures, lacks any significant observable role in the construction of the non-bonded contacts or assembly of *R* group side chains outside of weak van der Walls interactions.

Quasiracemates. The quasiracemic structures presented in this study include pairs of quasienantiomers where at least one of the

components contains an *R* group with a sulfur atom. Five of these amino acid components involve straight-chain *R* groups (Nva, Mecyst, Nle, Met, and Etcyst), while one entry, leucine, exists as a branched-chain side group. By pairing the L and D forms of these quasienantiomers with oxalic acid, eight new quasiracemic crystal structures were prepared (Fig. 1).

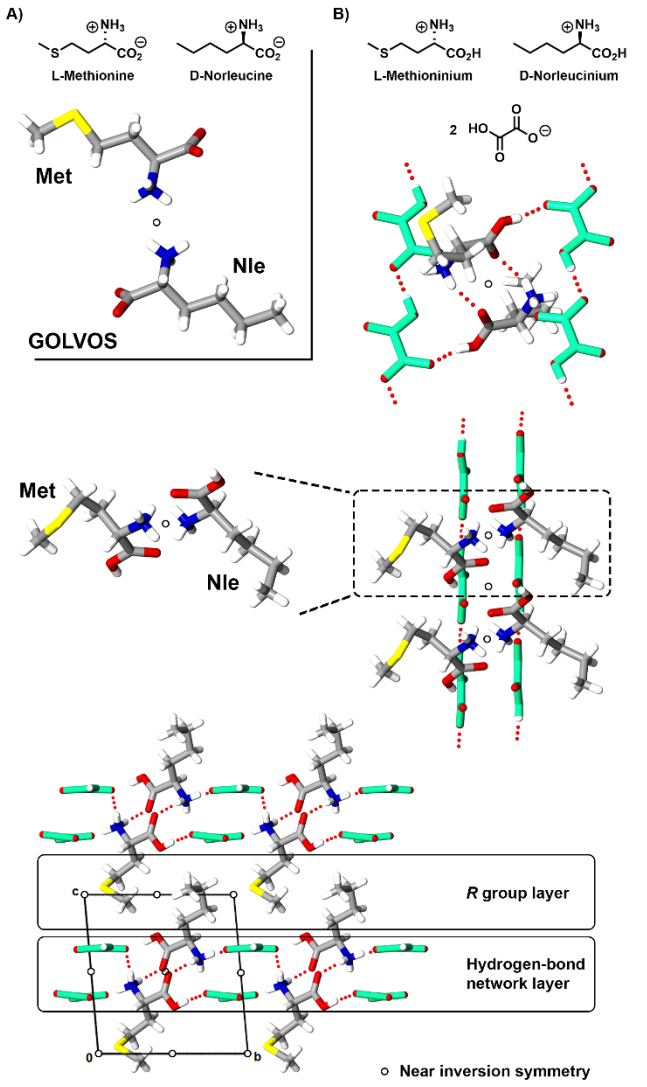

Fig. 3 Crystal structure views of quasiracemates L-Met·D-Nle (GOLVOS) and the hydrogen oxalate form (L-Met·D-Nle·2HOx).

Our involvement with sulfur-containing amino acid hydrogen oxalate quasiracemates initially targeted the Met·Nle·2HOx system. Because the selection of this quasiracemate included the use of two quasienantiomeric isosteres that differ only by -CH_2- and -S- substitutions, we anticipated this strategy would promote quasiracemate formation even in the presence of a secondary coformer molecule such as oxalic acid. This case study proved successful by providing a valuable opportunity to uncover the complex structural architectures present in the sulfur-based amino acid hydrogen oxalate quasiracemates. Fig. 3A depicts the crystal structure of the non-HOx and hydrogen oxalate forms of the L-Met·D-Nle system. The previously reported non-HOx form deposited in the *CCDC-Cambridge Structural Database*[32] (CSD, version 5.42, update 2), refcode GOLVOS[3]) crystallizes in chiral space group $P2_1$. Due to the location of the amino acid quasienantiomers in the crystal and the near inversion symmetry that relates these components, this structure closely mimics centrosymmetric space group $P2_1/c$ and the inversion symmetry elements found in the crystal packing of DL-Nle[33] and DL-Met[34,35] (Fig. 3A). The structure of the related hydrogen oxalate equivalent is noticeably different (Fig. 3B, space group *P*1). In this case, the quasienantiomers and hydrogen oxalate moieties align with Type I crystal packing *via* a complex network of hydrogen bonds that follows the structural patterns described for the analogous racemic systems.

The unusual feature of the Met·Nle·2HOx structure is that the spatial alignment of the central core of the amino acids – *i.e.,* H_3N^+-CH(CH_2)-CO_2H – closely mimics near inversion symmetry, while the remaining pendant *R* groups deviate significantly from this symmetry operator. This observation is significant and contrasts the generally accepted notion that quasiracemates nearly always crystallize with near inversion relationships between the quasienantiomeric components; structural patterns that strongly mimic the rigorously centrosymmetric packing found in the racemic counterparts. This correlation has been reported extensively, with many studies citing nearly indistinguishable or very similar crystallographic data sets, unit cell parameters, and crystal packing patterns for related racemates and quasiracemates.[36,37] To give some perspective to this phenomenon, the only outlier to the approximate centrosymmetric nature of quasiracemate structures we are aware of is a recent reinvestigation of Pasteur's 1850's tartaramide/malamide quasiracemates where the components organized with pseudo glide-plane symmetry[38]. For the Met·Nle·2HOx structure, the dimeric stacks of hydrogen oxalate chains and central amino acid framework strongly correlate to inversion symmetry and space group $P\bar{1}$. However, the spatial arrangement of the *R* groups of the norleucinium and methioninium components deviate significantly from this inversion symmetry relationship or other rational symmetry operators (crystallographic or noncrystallographic symmetry). This deviation from centrosymmetry appears to be more pronounced near the tail end of the amino acid chains.

Motivated by our experience with carbon-based amino acids and the structure of L-Met·D-Nle·2HOx, this study then focused on examining the structural effects and use of other sulfur-containing amino acids as building blocks in the construction of quasiracemic materials. Does the structural variance observed in L-Met·D-Nle·2HOx transfer to other sulfur-containing amino acid quasiracemates? In addition to pursuing this question *via* crystallographic investigations, this effort also quantitatively assessed the conformational similarity of the quasienantiomeric pair (χ_{RMS}) and how far these components differ from inversion symmetry (C_i). As previously described[1], Avnir's *Continuous Symmetry Measures*[39-41] (CSM) and the *Structure Overlay* utility in *CCDC-Mercury*[42] provided practical tools for these evaluations. Several structural modifications were applied before determining these numerical parameters.

Because both structural tools require symmetric sequences with shared atom connectivities for each quasiracemic pair, atoms that lacked an equivalent partner were omitted from the assessments. Because *CSM* distinguishes between atom types, sulfur atoms were replaced with carbon, and for the *CCDC-Mercury* studies, the stereochemistry of the quasienantiomeric pair was matched by inverting one of the amino acid components. Hydrogen atoms were also removed to streamline the process.

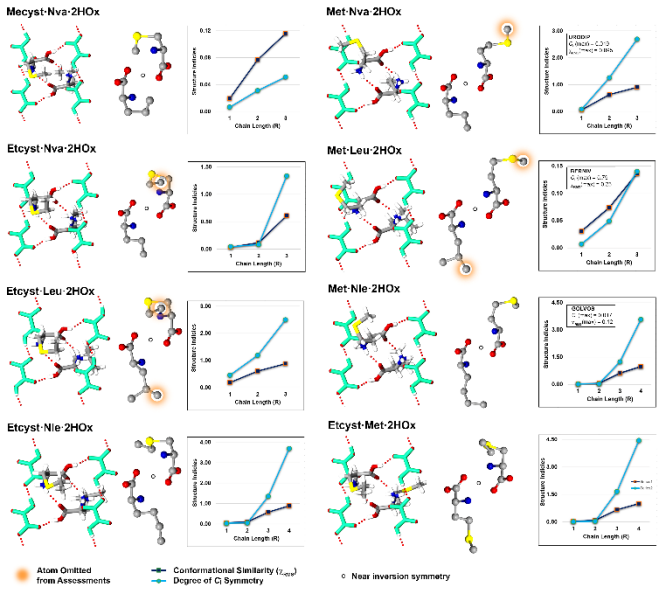

Fig 4. Eight amino acid hydrogen oxalate quasiracemates showing the crystal structures projected down the *c* axis, an isolated view of the quasiracemic pair showing molecular conformations (hydrogen atoms omitted for clarity), and a plot of the structure indices (χ_{RMS} and C_i) for each quasienantiomeric pair.

**** this figure should be included in the manuscript as a full-page figure (remove this comment)**

Fig. 4 shows the crystal structures of the eight amino acid hydrogen oxalate quasiracemates where the *R* group of at least one of the components contains sulfur. This homologous family displays a high degree of isostructurality with each crystallizing in space group *P*1 or $P2_1$, imitating $P\bar{1}$ or $P2_1/c$ due to the presence of near inversion symmetry, and exhibiting Type I hydrogen bond motifs. An interesting aspect of these closely related structures is the significant variability of *R* group spatial orientation within a quasienantiomeric pair and between these quasiracemic systems. A good indicator of this contrast can be found from visual inspection of the four structures containing the methioninium moiety (*i.e.,* L-Met·D-Nva·2HOx. L-Met·D-Leu·2HOx, L-Met·D-Nle·2HOx, L-Etcyst·D-Met·2HOx) (Fig. 4, right). Though the common feature is a methioninium component, various group geometries for the $-CH_2CH_2SCH_3$ fragment are present in these structures. This variation also exists when extending this qualitative assessment to all amino acids containing *R* groups with a straight chain of four atoms (*i.e.,* Met, Nle, and Etcyst). This assessment produced ten molecules from seven crystal structures with an equally drastic variation in *R* group conformations.

Since the core amino acid framework (H_3N^+-$CH(CH_2)$-CO_2H) associated with these structures visually indicates a greater extent of conformational similarity and symmetry alignment than the pendant *R* groups, we wondered how these measures correlate as a function of *R* group chain length. These strategies were recently applied to aliphatic amino acid hydrogen oxalates, where the molecular conformations (χ_{RMS}) and degree of inversion (C_i) between the two quasienantiomers closely correlate.[1] Inspection of the L-Met·D-Nle·2HOx example (Figs. 3B and 4) shows that the χ_{RMS} and C_i values associated with the core amino acid fragment are relatively small, indicating a strong conformational similarity and symmetry relationship of these groups. By contrast, including the *R* groups in the calculations significantly increased in the χ_{RMS} and C_i parameters. This increase correlates directly to chain length (χ_{RMS} (1-4) = 0.03, 0.06, 0.62, 0.99 and C_i (1-4) = 0.02, 0.04, 1.25, 3.58) and is consistent with the initial observation that the tail ends of the amino acids appear to deviate considerably from true inversion symmetry. However, when this approach was applied to the non-oxalate L-Met·D-Nle structure (GOLVOS), the results indicate components with similar geometries (χ_{RMS}(max) = 0.12) and an alignment that approximates near inversion symmetry (C_i(max) = 0.017). This trend is largely true for the other methionine amino acid quasiracemic structures provided in the literature [L-Val·D-Met (BERQIY), χ_{RMS}(max) = 0.05, C_i(max) = 0.07; L-Nva·D-Met (URODIP), χ_{RMS}(max) = 0.09, C_i(max), 0.02; L-Ile·D-Met (FITLID), χ_{RMS}(max) = 0.12, C_i(max) = 0.15; L-Leu·D-Met (BERNIV), χ_{RMS}(max) = 0.25, C_i(max) = 0.79; L-Phe·D-Met (POVYOP) χ_{RMS}(max) = 0.71 C_i(max) = 1.17]. An additional methionine quasiracemate, L-Abu·D-Met (ANUPOQ, ANUPOQ01 polymorphs), likely contains components with at least appreciably different quasienantiomeric rotamers; however, considerable disorder present in the *R* groups prevented suitable assessment of the structure indices.

A somewhat unanticipated result from this study was that the asymmetry displayed in the initial model structure of L-Met·D-Nle·2HOx exists in many of the other quasiracemic structures (Fig. 4). Five of these additional structures display C_i values greater than 1 with the most significant divergence being the L-Mecyst·D-Nva·2HOx system with C_i = 4.44 (L-Etcyst·D-Nva·2HOx, C_i(max) = 1.33; L-Etcyst·D-Leu·2HOx, C_i(max) = 2.50; L-Etcyst·D-Nle·2HOx, C_i(max) = 3.69; L-Met·D-Nva·2HOx, C_i(max) = 2.69). The remaining two systems, L-Mecyst·D-Nva·2HOx and L-Met·D-Leu·2HOx, form crystal structures where, despite containing the sulfur-based methioninium and methyl cysteinium fragments, the quasienantiomers closely mimic inversion symmetry (L-Mecyst·D-Nva·2HOx, C_i(max) = 0.05 and L-Met·D-Leu·2HOx, C_i(max) = 0.14). The conformational similarity for these same systems range from χ_{RMS}(max) = 0.12 to 1.00. (L-Mecyst·D-Nva·2HOx, χ_{RMS}(max) = 0.12; L-Etcyst·D-Nva·2HOx, χ_{RMS}(max) = 0.61; L-Etcyst·D-Leu·2HOx, χ_{RMS}(max) = 0.88; L-Etcyst·D-Nle·2HOx, χ_{RMS}(max) = 0.89; L-Met·D-Nva·2HOx, χ_{RMS}(max) = 0.90; L-Met·D-Leu·2HOx, χ_{RMS}(max) = 0.14; L-Met·D-Nle·2HOx, χ_{RMS}(max) = 0.99; L-Etcyst·D-Met·2HOx, χ_{RMS}(max) = 1.00). The structures with the largest χ_{RMS}(max) values correlate proportionally to larger C_i(max) values, indicating that pairs of quasienantiomers with

different geometries are often accompanied by crystal motifs that deviate considerably from inversion symmetry. Also, we were unable to determine a relationship between the spatial variation of quasienantiomeric components and the extent of divergence from inversion symmetry; HOx quasiracemates constructed from isosteric (D-Met·L-Nle·HOx) and or non-isosteric (D-Nva·L-Etcyst·HOx) quasienantiomers seem equally capable of forming highly asymmetric crystal assemblies.

Molecular Conformations

The conformational differences present in several of the quasienantiomeric pairs are largely unprecedented in the field of small molecule quasiracemic materials. While these structures maintain dimeric stacks of consistently spaced hydrogen-bonded HOx strands, the alignment of amino acid quasienantiomers linked to these motifs is often decidedly non-centrosymmetric. This departure from near inversion symmetry is largely inconsistent with the idea that centrosymmetric packing of achiral or racemic compounds yields energetically favorable motifs, where this penchant for inversion symmetry alignment has been estimated at >90% for racemic compounds.[43,44] So then, what structural factors drive these sulfur-based quasiracemates as the exception to the rule? Perhaps one area that could provide insight into this issue is the crystal growth experiments of these systems. All X-ray quality samples were prepared by room temperature straightforward cocrystallization experiments (S1, †). Crystals retrieved for both the racemic and quasiracemic crystal growth studies often formed thin layered plates, with many samples directly suitable for crystallographic assessment. Those samples requiring additional recrystallizations did not follow a consistent pattern related to amino acid selection or tabulated χ_{RMS} and C_i values. In the case of one of the racemates, DL-Nle·HOx, a previous report described this sample forms significantly curved crystals accompanied by whole-molecule disorder of the norleucinium molecules.[1] While this result may suggest that the centrosymmetric packing found in the frustrated crystals of DL-Nle·HOx (space group $P\bar{1}$) lack ideal crystal organization, this singular event is insufficient to explain why combining conformationally dissimilar L and D quasienantiomers is favored in many of the quasiracemic systems. It is also worth mentioning that centrosymmetric alignment does not always offer the most energetically favorable arrangement of molecules, as evidenced by the 8-9.5% of racemates that adopt non-centrosymmetric space groups[43,44]. While one could argue that the alignment of the *R* groups of the current quasiracemates closely resembles these non-centrosymmetric racemates, this is likely not the case since the components of the DL-amino acid·HOx structures assemble into rigorously centrosymmetric structures.

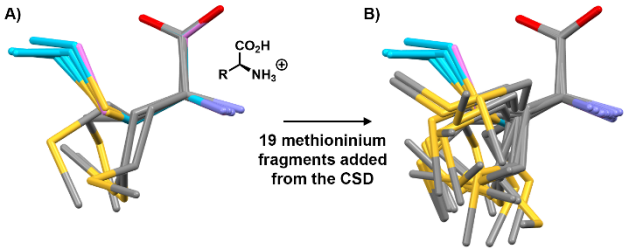

Fig. 5 CCDC-Mercury structure overlay plots of A) ten sulfur-containing amino acids [methyl cysteinium (1 entry, violet), ethyl cysteinium (4 entries, light blue), and methioninium (5 entries, cpk colors)] from the current study and B) the addition of nineteen methioninium CSD structures. H atoms were omitted for clarity.

To further understand the structural differences observed in the quasiracemate structures, this study examined the molecular conformations of each sulfur-containing amino acid component. The crystal structures were processed by isolating the desired molecule, matching the handedness of the quasienantiomers (L), and overlaying the set of compounds using the structure overlay function found in *CCDC-Mercury* (Fig 5A). The ten amino acids with sulfide groups, C-S-C, found in these eight crystal structures included methyl cysteinium (1 entry), ethyl cysteinium (4 entries), and methioninium (5 entries). Inspection of the overlay plot depicted in Fig. 5A shows that the core H_3N^+-CH(CH_2)-CO_2H fragments in these structures closely match, with DL-Met·HOx as the exception where the carboxyl hydrogen atom is located on the other carboxyl oxygen atom. The fixed geometry of the carboxyl group seems uniquely linked to the protonated forms of these amino acids since the non-hydrogen methionine oxalate structure, DL-Met, exists in two polymorphic forms (DLMETA07 and DLMETA08), with both displaying a variation in -CO_2^- group geometry. This same overlay data also shows that the *R* groups of the methyl and ethyl cysteinium molecules take on well-defined orientations. By contrast, the five methioninium entries from this study crystallize with a range of *R* group rotamers. A search of the CSD for structures containing each of the three sulfur-based amino acid fragments resulted in 19 additional methioninium molecules (Figure 5B). These entries serve to confirm that the core molecular framework of methioninium is retained, while the pendant side-chains take on a wide variety of conformations. Four of these CSD entries exist with two symmetry-independent methioninium molecules (LUDHEX, REMSUX, VICZEN, and YIGMEI). In these cases, the spatial arrangement of the -$CH_2CH_2SCH_3$ groups for each set of methioninium molecules from the same structure is significantly different, further supporting the observed structural pattern that methioninium readily takes on a range of conformations even within a given structure.

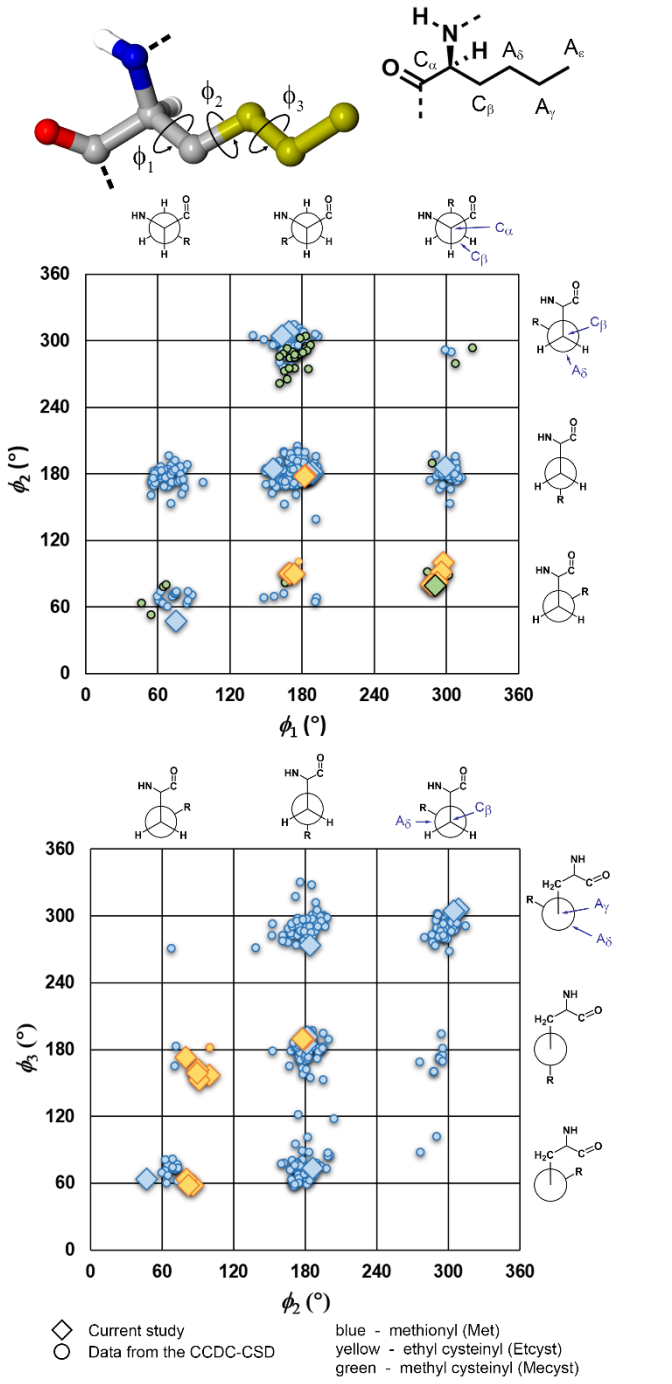

Fig. 6 Dihedral angle distributions from crystal structures containing methionyl, methyl cysteinyl, and ethyl cysteinyl fragments.

The conformational preferences identified for the sulfur-containing amino acids were further examined using the structural data from this study and the extant literature. A CSD search for the methionyl, methyl cysteinyl, ethyl cysteinyl fragments retrieved 292, 35, and 12 entries, respectively. Dihedral angle (ϕ) distributions are presented in Fig. 6 as 2D plots. As expected due to steric repulsions, the staggered geometries of these species effectively separate the data into roughly three rotamer bins ($0° < \phi < 120$, $120° < \phi < 240$, and $240° < \phi < 360$) for each dihedral angle.

Both gas phase calculations[45] and a search of high-resolution protein crystallographic data[46] have provided essential insight to preferred methionine geometries. These studies showed that the flexibility of the methionyl *R* group leads to a variety of low-energy geometries distributed over several conformers. And that the interplay of hyperconjugative effects and steric repulsion contributes to geometry stabilizations. As an example of the structural intricacies of these factors, the most destabilized conformers by steric hindrance are also the most stabilized species by hyperconjugation. Achieving a balance of these interrelated complex structural properties and others results in local/global energy minima corresponding in preferred geometries. When considering these factors in light of crystal packing forces, it is not surprising that multiple rotamers not only exist but compete during molecular assembly of the cocrystallization process. The populations of various rotamers of methionine and its analogues have been identified. Consistent with the computational[45] and protein crystal structures[46], the predominant rotamer displays ϕ_1 ~ 180°, where the *R* group is *anti* to the -CO_2R group. Furthermore, previous use of the protein and small molecule crystal structure databases provided additional evidence of preferred Met conformer orientations with ϕ_2 ~ 180° and ϕ_3 ~ 60° or 300°.[46] The geometry of the pendant methionine methyl group in these instances is inconsistent with that identified for Nle where ϕ_3 ~ 180°. Applying this information to the Met·Nle case study in this report revealed non-idealized component geometries for both Met·Nle (GOLVOS) [(Met) ϕ_1 = 187.6°, ϕ_2 ~ 180.6°, ϕ_3 = 189.2 and (Nle) ϕ_1 = 172.4°, ϕ_2 = 177.0°, ϕ_3 = 174.3] and Met·Nle·2HOx [(Met) ϕ_1 = 163.5°, ϕ_2 = 304.2°, ϕ_3 = 304.3 and (Nle) ϕ_1 = 187.7°, ϕ_2 = 168.3°, ϕ_3 = 66.2°]. For Met·Nle (GOLVOS), the components exhibit complementary anti-staggered geometries with ϕ_3(Met) skewed by ~60° from the idealized structure. A similar departure in dihedral angle is observed in the structure of Met·Nle·2HOx for both the ϕ_2(Met) and ϕ_3(Nle) parameters.

Inspection of the two panels of dihedral angle distributions provided in Fig. 6 shows the spatial variation of the amino acid *R* groups. For the methionyl fragment, the data (292 entries) clusters at ϕ_1 ~ 180°, with 67% of the total population, at ϕ_2 ~ 180 (73%) and ϕ_3 ~ 60° (30%), 180° (27%), and 300° (43%). The methyl cysteinyl structures (35 entries) cluster at ϕ_1 ~ 180° (68%) and ϕ_2 ~ 300° (71%) and those with the ethyl cysteinyl framework (12 entries) cluster at ϕ_1 ~ 180° (50%) and 300° (50%), ϕ_2 ~ 60° (83%), and ϕ_3 ~ 180° (67%). While each of the three fragment types shows ϕ_1 dihedral angle distributions favoring ~ 180°, the ϕ_2 values corresponding to the methyl and ethyl cysteinyl fragments, species that only differ by a methyl group, vary by ~ 120°. Of additional importance to the current discussion is that the nine methionyl fragments reported in this study show similar patterns in ϕ_1, ϕ_2, and ϕ_3 distributions to those previously reported for entries from protein crystal structures.

An important aspect of this study relates to the lack of complementary conformations observed for the quasiracemate components. Fig. 5 shows that the ethyl cysteinium fragment prefers ϕ_1 ~ 300°, ϕ_2 ~ 100°, and ϕ_3 ~ 160°, while the

methioninium components from our study take on a variety of conformations. Also, the rotamer of methyl cysteinium in Fig. 5 is similar to ethyl cysteinium. A CSD search for methionyl, methyl cysteinyl, and ethyl cysteinyl fragments increases the number of dataset entries and shows corresponding dihedral angle distributions (Fig. 6). Though there is some overlap with the methionyl and ethyl cysteinyl entries, the data for these fragments cluster in different regions of the two panels. This information provides evidence that the position of the sulfur atom along the side chain (*i.e.,* methionyl and ethyl cysteinyl) results in different energetically favorable conformations and that this variance in rotamer structures is accommodated in the crystal. Additionally, because the *R* groups reside in bilayer regions that lack strong non-bonded contacts, the crystal packing forces needed for near inversion symmetry alignment are likely minimal or at least less than what is need to overcome the preferred *R* group conformations.

New Additions to Amino Acid Quasiracemates

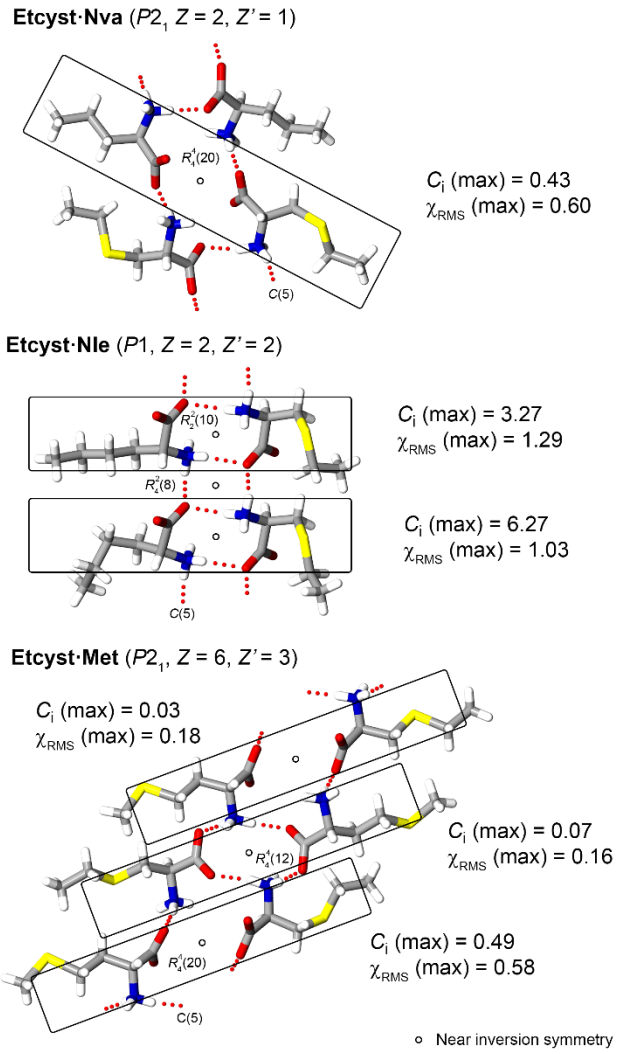

Fig. 7 Crystal structure representations and structure indices (χ_{RMS} and C_i) for Etcyst·Nva, Etcyst·Nle, and Etcyst·Met.

During the course of these studies, several non-hydrogen oxalate quasiracemates containing ethyl cysteine (*i.e.,* Etcyst·Nva, Etcyst·Nle, and Etcyst·Met) were prepared and crystal structures determined (Fig. 7). The hydrogen-bond patterns for these systems show components linked by a series of $CO_2^-\cdots H_3N^+$ interactions producing *C*(5) motifs and other cyclic hydrogen-bond patterns (*i.e.,* 8, 10, 12, or 20-membered hydrogen-bonded rings). Similar to the amino acid·HOx structures, these quasiracemic systems align components into bilayer structures with two distinct regions of closely aligned *R* groups and 2D hydrogen-bonded networks involving the core H_3N^+-CH(CH_2)-CO_2^- amino acid framework. While ethyl cysteine provided the initial common feature for these structures, several distinguishing structural traits exist, such as the variance in the number of symmetry unique molecules (*Z*′) and distinct rotamers, even when considering the multiple occurrences of the ethyl cysteine fragment. This lack of conformational similarity is all the more striking when considering the structures Etcyst·Nle, Etcyst·Met, and Met·Nle where the *R* group chain length is the same, and the isosteric components only differ by the placement of sulfur atoms or the S/CH_2 substitutions.

Calculation of the structure indices χ_{RMS} and C_i for these three systems used pairs of quasienatiomers located in close proximity and linked by hydrogen bonds. For the EtCyst·Nva and EtCyst·NMet structures, the χ_{RMS} and C_i values of < 0.6 indicate the components show a high degree of conformational similarity and align with near inversion symmetry. Cocrystallizing Etcyst with Met provides quite a different structural landscape where the two sets of symmetry-independent molecules result in sizable χ_{RMS} (1.29 and 1.03) and C_i (3.27 and 6.27) values. The importance of this structure is that it offers the only example of a non-HOx amino acid quasiracemate with crystal alignment that deviates significantly from the non-centrosymmetric crystal packing.

Conclusions

A new set of amino acid hydrogen oxalate quasiracemates that differ in their side groups have been designed, synthesized, and characterized. Crystal structures of these systems show the side groups align with decidedly non-centrosymmetric packing, while the core amino acid frameworks (*i.e.,* H_3N^+-CH(CH_2)-CO_2^-) and hydrogen oxalate column motifs closely mimic this symmetry element. Further quantitative assessment of these structures looked at the conformational similarity of each quasienantiomeric pair (χ_{RMS}) and how far these amino acid components differ from inversion symmetry (C_i). The results show that six of these systems display C_i values > 1.3 accompanied by different quasienantiomeric geometries (χ_{RMS} > 0.6). Large χ_{RMS} and C_i values were also determined for the non-HOx system L-Etcyst·D-Met. The observed structural deviation from inversion symmetry lacks any apparent correlation to the length, composition, or geometry of the *R* group. To the best of our knowledge, this study represents a new addition to quasiracemic materials where the components deviate significantly from near inversion or glide-plane

symmetry. Because the motifs found in these amino acid·HOx systems provide a unique detour from all recognized crystalline quasiracemates, these results promise to open yet further perspectives in quasiracemate chemistry. Additional work to uncover the structural features responsible for the asymmetric alignment of the amino acid side groups is ongoing in our laboratory, and the results will be disclosed in due course.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was generously supported by the National Science Foundation (DMR1904651 and CHE1827313) and Whitworth University. We also thank Prof. A. L. Rheingold (Univ. of California San Diego) for key crystallographic contributions to this study.

Notes and references

1 R. G. Wells, K. D. Sahlstrom, F. I. Ekelem and K. A. Wheeler, *CrystEngComm*, 20##, **##**, ### – ###.
2 C. H. Görbitz, K. Rissanen, A. Valkonen and A. Husabø, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun*., 2009, **C65**, o267 – o272. L-Phe·D-Abu (CSD ref. code: POVYEF), L-Phe·D-Nva (POVYIJ), L-Phe·D-Met (POVYOP), L-Phe·D-Leu (POVYUV), L-Phe·D-Ile (POVZAC), L-Phe·D-aIlo·Ile (POXGAL).
3 B. Dalhus and C. H. Görbitz, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun*., 1999, **C55**, 1105 – 1112. D-Nle·L-Nva (GOLVIM), D-Nle·L-Met (GOLVOS), D-Nle·L-Val (GOLVUY), D-Nle·L-aIle (GOLWAF), D-Nle·L-Leu (GOLWEJ).
4 Dalhus and C. H. Görbitz, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.*, 1999, **C55**, 1547 – 1555. L-Val·D-Abu (BERQAQ), L-Val·D-Nva (BERQEU), L-Val·D-Met (BERQIY), L-Leu·D-Abu (BERNAN), L-Leu·D-Nva (BERNER), L-Leu·D-Met (BERNIV), L-Leu·D-Val (BERPET).
5 B. Dalhus and C. H. Görbitz, *Acta Crystallogr. Sect. B Struct. Sci*., 1999, **B55**, 424 – 431. L-Ile·D-Ala (FITHIZ), L-Ile·D-Abu (FITJAT), L-Ile·D-Nva (FITJEX), L-Ile·D-Nle (FITLEZ), L-Ile·D-Met (FITLID), L-Ile·D-Val (FITMEA), L-Ile·D-Leu (FITNIF).
6 M. M. H. Smets, E. Kalkman, P. Tinnemans, A. M. Krieger, H. Meeks and H. M. Cuppen, *CrystEngComm*, 2017, **19**, 5604 – 5610. D-Abu·L-Nva (VEDZUC, VEDZUC01).
7 C. H. Görbitz, D. S. Wragg, I. M. B. Bakke, C. Fleischer, G. Grønnevik, M. Mykland, Y. Park, K. W. Trovik, H. Serigstad and B. E. V. Sundsli, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun*., 2016, **C72**, 536 – 543. L-Abu·D-Met (ANUPOQ, ANUPOQ01 polymorphs).
8 C. H. Görbitz and P. Karen, *J. Phys. Chem. B*, 2015, **119**, 4975 – 4984. L-Nle·D-Met (VUQNAY).
9 C. H. Görbitz, B. Dalhus and G. M. Day, *Phys. Chem. Chem. Phys.*, 2010, **12**, 8466 – 8477. L-Abu·D-Nle (URODOV), L-aIlo Ile·D-Leu (URODEL), L-Nva·D-Met (URODIP).
10 G. S. Prasad and M. Vijayan, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun*., 1991, C47, 2603 – 2606. L-Phe·D-Val (SONCED).
11 A. Fredga, *Bull. Soc. Chim. Fr.*, 1973, **1**, 173 – 182.
12 A. Fredga, *Tetrahedron*, 1960, **8**, 126 – 144.
13 M. S. Nandhini, R. V. Krishnakumar and S. Natarajan, *Acta Crystallogr., Sect. E: Struct. Rep. Online*, 2001, **57**, o633 – o635. L-Ala·HOx (YEJYIV).
14 K. Rajagopal, R. V. Krishnakumar, M. S. Nandhini, R. Malathi, S. S. Rajan and S. Natarajan, *Acta Crystallogr., Sect. E: Struct. Rep. Online*, 2003, **59**, o878 – o880. L-Leu·HOx (HAGZUL).
15 V. S. Minkov and E. V. Boldyreva, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.*, 2008, **64**, o344 – o348. L-Cyst·HOx (LOCLOF).
16 R. P. Sukiasyan, H. A. Karapetyan and A. M. Petrosyan, *J. Mol. Struct*., 2008, **888**, 230 – 237. L-Lys HOx (MOHDIX).
17 N. R. Chandra, M. M. Prabu, J. Venkatraman, S. Suresh and M. Vijayan, *Acta Crystallogr., Sect. B: Struct. Sci*., 1998, **54**, 257 – 263. L-Arg·HOx (NOSXEY).
18 M. M. Prabu, H. G. Nagendra, S. Suresh and M. Vijayan, J. *Biomol. Struct. Dyn*., 1996, **14**, 387 – 392. L-His·HOx (RARXOX).
19 D. Braga, L. Chelazzi, I. Ciabatti and F. Grepioni, *New J. Chem*., 2013, **37**, 97 – 104. L-Ser·HOx (XENXOF).
20 O. Bakke and A. Mostad, *Acta Chem. Scand*., 1980, **34**, 559 – 570. L-Trp·HOx (TRYPTB).
21 B. A. Zakharov and E. V. Boldyreva, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun*., 2011, **67**, o47 – o51. DL-Ala·HOx (IMEGIR).
22 M. M. Prabu, H. G. Nagendra, S. Suresh, M. Vijayan, *J. Biomol. Struct. Dyn.*, 1996, **14**, 387 – 392. DL-His·HOx (RARXUD).
23 M. S. Nandhini, R. V. Krishnakumar, R. Malathi, S. S. Rajan and S. Natarajan, *Acta Crystallogr., Sect. E: Struct. Rep. Online*, 2001, **57**, o769 – o771. DL-Thr·HOx (REPFEX).
24 K. Manoj, H. Takahashi, Y. Morita, R. G. Gonnade, S. Iwama, H. Tsue and R. Tamura, *Chirality*, 2015, **27**, 405 – 410. DL-Leu·HOx (WIPQOE).
25 J. Venkatraman, M. M. Prabu and M. Vijayan, *J. Pept. Res*., 1997, **50**, 77 – 87. DL-Lys·HOx (XOXGUM).
26 V. S. Minkov and E. V. Boldyreva, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun*. 2009, **65**, o245 – o247. DL-Cys·HOx (BOWKOO).
27 V. S. Minkov, E. V. Boldyreva, T. N. Drebushchak and C. H. Gorbitz, *CrystEngComm*, 2012, **14**, 5943 – 5954. DL-Cys·HOx (BOWKOO01).
28 N. R. Chandra, M. M. Prabu, J. Venkatraman, S. Suresh and M. Vijayan, *Acta Crystallogr., Sect. B: Struct. Sci*., 1998, **54**, 257 – 263. DL-Arg·HOx (NOSXAU).
29 A. McL. Mathieson, *Acta Crystallogr.*, 1953, **6**, 399 – 403. (DLNLUA).
30 J. Bernstein, R. E. Davis, L. Shimoni and N.-L. Chang, *Angew. Chem. Int. Ed.*, 1995, **34**, 1555 – 1573.
31 M. C. Etter, *Acct. Chem. Res.*, 1990, **23**, 120 – 126.
32 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr. Sect. B Struct. Sci*., 2016, **B72**, 171 – 179.
33 S. J. Coles, T. Gelbrich, U. J. Griesser, M. B. Hursthouse, M. Pitak and T. Threlfall, *Cryst. Growth Des*., 2009, **9**, 4610 – 4612 (DLNLUA03).
34 A. McL. Mathieson, *Acta Crystallogr*. 1952, **5**, 332 – 341. (DLMETA).
35 M. M. H. Smets, S. J. T. Brugman, E. R. H. van Eck, P. Tinnemans, H. Meekes and H. M.Cuppena, *CrystEngComm*, 2016, **18**, 9363 – 9373. (DLMETA10)
36 D. F. Kreitler, D. E Mortenson, K. T. Forest and S. H. Gellman, *J. Am. Chem. Soc*., 2016, **138**, 6498 – 6505.
37 J. M. Spaniol and K. A. Wheeler, *RSC Adv.*, 2016, **6**, 64921 – 64929.
38 E. M. Pinter, L. S. Cantrell, G. M. Day and K. A. Wheeler, *CrystEngComm*, 2018, **20**, 4213-4220.
39 P. Alemany, D. Casanova, S. Alvarez, C. Dryzun and David Avnir, "Continuous Symmetry Measures: A New Tool in Quantum Chemistry" Reviews in Computational Chemistry, Volume 30, Abby L. Parrill and Kenneth B. Lipkowitz, Editors, Chapter 7, p. 289 – 352, Wiley, 2017.
40 M. Pinsky, C. Dryzun, D. Casanova, P. Alemany and D. Avnir, *J. Comput. Chem*., 2008, **29**, 2712 – 2721.

41 H. Zabrodsky, S. Peleg and D. Avnir, *J. Am. Chem. Soc.*, 1992, **114**, 7843 – 7851.
42 C. F. Macrae, I. Sovago, S. J. Cottrell, P. T. A. Galek, P. McCabe, E. Pidcock, M. Platings, G. P. Shields, J. S. Stevens, M. Towler and P. A. Wood, *J. Appl. Cryst.*, 2020, **53**, 226 – 235.
43 T. Rekis, *Acta Crystallogr. Sect. B Struct. Sci.*, 2020, **B76**, 307 – 315.
44 S. P. Kelley, L. Fabian and C. P. Brock, *Acta Crystallogr. Sect. B Struct. Sci.*, 2011, **B67**, 79 – 93.
45 W. G. D. P. Silva, C. B. Braga and R. Rittner, *Beilstein J. Org. Chem.*, 2017, **13**, 925 – 937.
46 A. Virrueta, C. S. O'Hern and L. Regan, *Proteins*, 2016, **84**, 900 – 911.