Acta Crystallographica Section C
Crystal Structure Communications
ISSN 0108-2701

A new 'hydrogen-bond rule' applied to the structure of L-seryl-L-alanine and pairs of dipeptide retroanalogues

Carl Henrik Görbitz

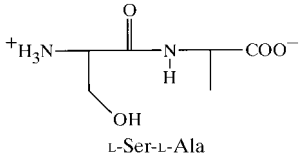
Department of Chemistry, University of Oslo, PO Box 1033 Blindern, N-0315 Oslo, Norway
Correspondence e-mail: c.h.gorbitz@kjemi.uio.no

Received 8 November 1999
Accepted 26 January 2000

A new 'rule' for the association of hydrogen-bond donors and acceptors in crystal structures is presented. It implies that ranks are assigned to each donor and each acceptor (1 is best, 2 is next best *etc*.), and that hydrogen bonds should be formed between donors and acceptors in rank order. L-Ser-L-Ala, $C_6H_{12}N_2O_4$, is used together with its retroanalogue, L-Ala-L-Ser, and three other pairs of dipeptide retroanalogues to illustrate this rule and the reasons why it may not always be followed.

Comment

One of the empirical 'Hydrogen-Bond Rules' (Etter, 1990) states that 'the best proton donors and acceptors remaining after intramolecular hydrogen-bond formation form intermolecular hydrogen bonds to one another'. This rule may tentatively be expanded to include all acceptors and donors in a new hydrogen-bond Network Rule (NR) as follows: 'after intramolecular hydrogen-bond formation, proton donors and acceptors associate in rank order'. In principle, any structure can be analysed with reference to this rule. Two previous papers (Görbitz & Backe, 1996; Görbitz, 1999*a*) have demonstrated that studies of dipeptide retroanalogues are quite useful. The present work concludes this investigation.



L-Ser-L-Ala

The structure of L-Ser-L-Ala is shown in Fig. 1. The structure of its retroanalogue, L-Ala-L-Ser, was presented by Jones *et al.* (1978*a*). For both compounds, the three best hydrogen-bond donors, with ranks 1, 2 and 3, are the three amino N—H atoms, while the carboxylate group has acceptor ranks from 1 to 4. As far as the two additional donors (>N—H and –CH_2–OH) are concerned, it is not obvious which should be assigned rank 4 and which rank 5. This is also true for assigning ranks to the two acceptors >C═O and –CH_2–OH, although judging by statistical values for the donor···O distances with carbonyl and water acceptors (Görbitz, 1989), one can tentatively assign rank 5 to the hydroxyl group and rank 6 to >C═O. With this set of ranks, we find that hydrogen bonding in the L-Ala-L-Ser structure (Fig. 2*b*; Jones *et al.*, 1978*a*) strictly follows the extended hydrogen-bond rule, as does L-Ser-Gly (Jones *et al.*, 1978*b*), in which precisely the same types of interactions occur. L-Ser-L-Ala (Fig. 2*a*) represents only a modest deviation from this pattern, in that one of the amino NH protons is accepted by the L-Ser side-chain hydroxyl group rather than by the main chain carboxylate group. As discussed previously, however, hydrogen bonding in Gly-L-Ser (Görbitz, 1999*a*) is quite different. It is interesting to see how the introduction of a small hydrophobic group [H (Gly) → methyl (Ala)] gives more similar hydrogen-bond interactions within the retroanalogue pair. It can be seen from Fig. 2 how these methyl groups generate small hydrophobic columns along the shortest axis in each structure. This is a common motif for the aggregation of hydrophobic groups in the crystal structures of peptides (Görbitz & Etter, 1992).

Dipeptide structures have previously (Görbitz, 1999*a*) been retrieved from the Cambridge Structural Database (Allen & Kennard, 1993) and divided into three categories (counting Gly as a hydrophilic residue): (*A*) hydrophilic structures with abundant hydrogen bonding, D_x is typically 1.40–1.60 Mg m^{-3}; (*B*) dipeptides with one hydrophobic residue, D_x is in the range 1.25–1.40 Mg m^{-3} (1.44 for L-Ser-L-Ala represents an extreme value); (*C*) strictly hydrophobic dipeptides, D_x is in the range 1.05–1.20 Mg m^{-3}.

The L-Ala-L-Ser/L-Ser-L-Ala pair belongs to group *B*, as does the L-Val-L-Glu (Eggleston, 1984)/L-Glu-L-Val (Görbitz & Backe, 1996) pair. The Gly-L-Ser/L-Ser-Gly pair discussed above, on the other hand, belongs to group *A*. Furthermore, we have data for a fourth pair, L-Val-L-Ala (Görbitz & Gundersen, 1996) and L-Ala-L-Val (Görbitz, 2000), belonging to group *C*. It should be added that dipeptides (as well as other peptides) often include cocrystallized water or organic solvent molecules. Therefore, the structures of retroanalogues may not contain the same hydrogen-bond donors and acceptors [*e.g.* Gly-L-Asp·2H_2O (Eggleston & Hodgson, 1982)/L-Asp-Gly·H_2O (Eggleston *et al.*, 1984) and L-Ala-L-Leu·0.5H_2O (Görbitz, 1999*b*)/L-Leu-L-Ala·4H_2O (Görbitz, 1997)], and a

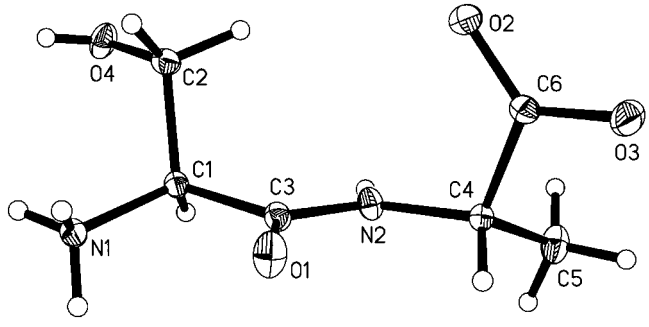
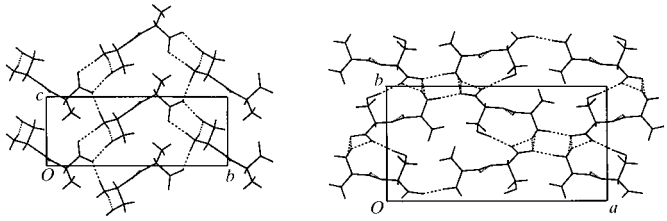


Figure 1
The asymmetric unit of L-Ser-L-Ala with the atomic numbering. Displacement ellipsoids are shown at the 50% probability level and H atoms are shown as spheres of arbitrary size.

direct comparison of hydrogen-bonding patterns is rendered less straightforward.

Some relevant data for the four known pairs of solvent-free retroanalogue pairs are given in Table 3. The most obvious observation is that an increasing number of hydrophobic groups in a molecule (meaning an unchanged or reduced number of hydrophilic groups) makes hydrogen bonding within a pair more similar, but it does not imply that the NR is followed more rigorously. This apparent contradiction may be explained by considering the problems associated with arranging three main chain carboxylate groups (as required by the NR) around each amino group when usually rather bulky hydrophobic side chains are present. In fact, only about one out of eight dipeptide structures displays three amino N—H$^{+}\cdots{}^{-}$O—C carboxylate hydrogen bonds (Görbitz, 1999*a*), usually when Gly (or less frequently Ala) is either an N-terminal or a C-terminal residue. We believe that deviations from the NR can, at least in part, be explained by such inherent steric constraints, combined with the need always to segregate hydrophobic groups into distinct regions of the crystal.

The experimental material discussed in this paper is limited, but the results should give at least an indication of the general trends for short linear peptides. It is clear that the NR is not very robust, as it is followed (completely or almost) by only three out of the eight structures discussed here. Nevertheless, the rule could be a useful tool in the analysis of two- and three-dimensional hydrogen-bond networks in crystal structures of a variety of organic compounds. Further results from such investigations would be most interesting.



(*a*) (*b*)

Figure 2
The unit cell and hydrogen-bond pattern for (*a*) L-Ser-L-Ala and (*b*) L-Ala-L-Ser (Jones *et al.*, 1978*a*). Views are along the 4.849 Å *a* axis and the 4.859 Å *c* axis, respectively.

Experimental

The title compound was obtained from Sigma. The specimen used for data collection was the only obvious single crystal resulting from a series of slow evaporation experiments with aqueous solutions of the dipeptide at room temperature.

Crystal data

$C_6H_{12}N_2O_4$
M_r = 176.18
Monoclinic, $P2_1$
a = 4.8488 (1) Å
b = 14.8294 (4) Å
c = 6.0228 (2) Å
β = 110.534 (1)°
V = 405.553 (19) Å^3
Z = 2
D_x = 1.443 Mg m^{-3}
Mo $K\alpha$ radiation
Cell parameters from 4637 reflections
θ = 3–35°
μ = 0.121 mm^{-1}
T = 153 (2) K
Block, colourless
0.42 × 0.35 × 0.17 mm

Data collection

Siemens SMART CCD diffractometer
Sets of exposures each taken over 0.6° ω rotation scans
Absorption correction: empirical (*SADABS*; Sheldrick, 1996)
T_{min} = 0.95, T_{max} = 0.98
6572 measured reflections
1837 independent reflections
1786 reflections with $I > 2\sigma(I)$
R_{int} = 0.033
θ_{max} = 34.97°
$h = -7 \rightarrow 7$
$k = -23 \rightarrow 23$
$l = -9 \rightarrow 9$
Intensity decay: none

Refinement

Refinement on F^2
$R[F^2 > 2\sigma(F^2)]$ = 0.057
$wR(F^2)$ = 0.127
S = 1.342
1837 reflections
120 parameters
H atoms treated by a mixture of independent and constrained refinement
$w = 1/[\sigma^2(F_o^2) + (0.0432P)^2 + 0.176P]$ where $P = (F_o^2 + 2F_c^2)/3$
$(\Delta/\sigma)_{max}$ = 0.002
$\Delta\rho_{max}$ = 0.42 e Å^{-3}
$\Delta\rho_{min}$ = −0.26 e Å^{-3}

Table 1
Selected geometric parameters (Å, °).

O1—C3	1.238 (3)	O3—C6	1.276 (3)
O2—C6	1.242 (3)	N1—C1	1.484 (3)
N1—C1—C3—N2	163.35 (17)	N2—C4—C6—O2	−10.1 (3)
C1—C3—N2—C4	172.41 (17)	N1—C1—C2—O4	−76.6 (2)
C3—N2—C4—C6	−80.4 (2)		

Table 2
Hydrogen-bonding geometry (Å, °).

D—H···A	D—H	H···A	D···A	D—H···A
N1—H1···O4^{i}	0.88	1.97	2.831 (3)	167
N1—H2···O3^{ii}	0.88	1.85	2.723 (3)	173
N1—H3···O2^{iii}	0.88	1.89	2.718 (3)	157
N2—H4···O1^{iv}	0.77	2.20	2.947 (2)	162
O4—H5···O3^{iii}	0.75 (4)	1.88 (4)	2.623 (3)	173 (4)

Symmetry codes: (i) $x-1, y, z$; (ii) $1-x, y-\frac{1}{2}, 2-z$; (iii) $1-x, y-\frac{1}{2}, 1-z$; (iv) $1+x, y, z$.

Table 3
Comparison of hydrogen-bond types in the structures of dipeptide retroanalogues.

Pair	D_x/Mg m^{-3}	Hydrogen-bond types	Follows NR	Reference†
Gly-L-Ser/ L-Ser-Gly	1.55/1.60	Rather different	No/yes	(*a*)/(*b*)
L-Ala-L-Ser/ L-Ser-L-Ala	1.42/1.44	Very similar	Yes/almost	(*c*)/(*d*)
L-Val-L-Glu/ L-Glu-L-Val	1.31/1.38	Similar	Partly/partly	(*e*)/(*f*)
L-Ala-L-Val/ L-Val-L-Ala	1.07/1.04	Similar*	Partly/partly	(*g*)/(*h*)

† (*a*) Görbitz (1999*a*), (*b*) Jones *et al.* (1978*a*), (*c*) Jones *et al.* (1978*b*), (*d*) present work, (*e*) Eggleston (1984), (*f*) Görbitz & Backe (1996), (*g*) Görbitz (2000), (*h*) Görbitz & Gundersen (1996); * = isomorphous.

The hydroxylic H atom was refined isotropically. Other peptide H atoms were placed geometrically and refined with constraints to keep all C—H/N—H distances and all C—C—H/C—N—H angles on one

C or N atom the same. U_{iso} values were $1.2U_{eq}$ of the carrier atom, or $1.5U_{eq}$ for hydroxyl, methyl and amino groups. Free rotation was permitted for amino and methyl groups.

Data collection: *SMART* (Siemens, 1995); cell refinement: *SAINT* (Bruker, 1997); data reduction: *SAINT*; program(s) used to solve structure: *SHELXTL* (Sheldrick, 1997); program(s) used to refine structure: *SHELXTL*; molecular graphics: *SHELXTL*; software used to prepare material for publication: *SHELXTL*.

The purchase of the Siemens SMART diffractometer was made possible through support from The Research Council of Norway (NFR).

Supplementary data for this paper are available from the IUCr electronic archives (Reference: DE1127). Services for accessing these data are described at the back of the journal.

References

Allen, F. H. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 31–37.
Bruker (1997). *SAINT*. Version 5.00. Bruker AXS Inc., Madison, Wisconsin, USA.
Eggleston, D. S. (1984). *Acta Cryst.* C**40**, 1250–1252.
Eggleston, D. S. & Hodgson, D. J. (1982). *Int. J. Peptide Protein Res.* **19**, 206–211.
Eggleston, D. S., Valente, E. J. & Hodgson, D. J. (1984). *Acta Cryst.* B**37**, 1428–1430.
Etter, M. C. (1990). *Acc. Chem. Res.* **23**, 120–126.
Görbitz, C. H. (1989). *Acta Cryst.* B**45**, 390–395.
Görbitz, C. H. (1997). *Acta Cryst.* C**53**, 736–739.
Görbitz, C. H. (1999*a*). *Acta Cryst.* B**55**, 1090–1098.
Görbitz, C. H. (1999*b*). *Acta Cryst.* C**55**, IUC9900149.
Görbitz, C. H. (2000). In preparation.
Görbitz, C. H. & Backe, P. H. (1996). *Acta Cryst.* B**52**, 999–1006.
Görbitz, C. H. & Etter, M. C. (1992). *Int. J. Peptide Protein Res.* **39**, 93–110.
Görbitz, C. H. & Gundersen, E. (1996). *Acta Cryst.* C**52**, 1764–1767.
Jones, P. G., Falvello, L. & Kennard, O. (1978*a*). *Acta Cryst.* B**34**, 1939–1942.
Jones, P. G., Falvello, L. & Kennard, O. (1978*b*). *Acta Cryst.* B**34**, 2379–2381.
Sheldrick, G. M. (1996). *SADABS*. University of Göttingen, Germany.
Sheldrick, G. M. (1997). *SHELXTL*. Version 5.10. Bruker AXS Inc., Madison, Wisconsin, USA.
Siemens (1995). *SMART*. Version 4.05. Siemens Analytical X-ray Instruments Inc., Madison, Wisconsin, USA.