Crystal structures of dipeptides: the head-to-tail story

Carl Henrik Görbitz

Department of Chemistry, University of Oslo, Norway

List of supplementary material:

Table 1S
Data set sorted by number of *C*(8) head-to-tail chains, *N*

REFCODE[a]	seq.[b]	stereo[c]	solvent/guest/anion	Description[d]
N = 0 (15)				
BEQJAJ	IW		dihydrate	**T5-S5****
BEQJEN	WV		dihydrate	**S5****
BUDXUT	PE		dihydrate	**T5****
CAXNUK	**RE**		dihydrate	no layer, *Z*' = 2
CELTAO10	YF		hydrate	**T5****
CIHNUC	YV		hydrate	**T5****
ETONIK	IF		dihydrate	**T5-S5**** = BEQJAJ
GASWEC	PK+		acetate	no layer
KIXCAW	YW		hydrate	**T5****
MOBYAD	VF		dihydrate	**T5-S5**** = BEQJAJ
RAVZEU	GH		dihydrate	**T5****
VEVGOS	**KE**		dihydrate	no layer
VUZBAT	**RD**		dihydrate	no layer
VUZBIB	YL		hydrate	**T5****
ZEFZAL10	YY		dihydrate	no layer
N = 1 (27)				
ALGLYL	AG		Li^+,Br^-,dihydrate	no layer
ASPGLY	DG		hydrate	**S4***
BEVXEF01	GD		dihydrate	no layer
CEFGOJ	YE		hydrate	**T5***
DIYZEQ01	**RD**		hydrate	no layer
GLLASP	GN			(**S5***)
GLTLYR10	GY		dihydrate	**T4**** helix
GLTRDH01	GW		dihydrate	**S5***
GLYGCA	GG		Ca^{2+}, $2Cl^-$	no layer
GLYGLB	GG		Li^+, Br^-	no layer
GLYTRE04	GT		dihydrate	anti = GLYGLY04
JADVAL	FP		hydrate	anti
KIXBOJ	SY		hydrate	**S4***
KIXBUP	WS		hydrate	straight chains
LACBAS	GH+		semisuccinate hydrate	**S4***
LPROHP20	PhP		hydrate	no layer
MAVTAE	nGfG		(S)-methyl lactate clathrate	anti, *Z*' = 2
MAVTIM	nGfG		(*S*)-ethyl pantolactone dimethanol clathr. hydrate	anti = MAVTAE, *Z*' = 2
MOBYEH	VF		trihydrate	complex, *Z*'= 4
POTPET02	GG		1.5 hydrate	**S4***
QIMBUJ	FA		dihydrate	screw chains
RAVMAC	HG+		chloride	straight chains
RAVMOQ	LA		tetrahydrate	**S5***
RAVZIY	LH		hydrate	**S5***
RAVZUK	**HE**			**S4***
TEJGAQ	GH+		chloride dihydrate	**S4***, very folded
YAGZOW	II		dihydrate	(**S5***)
N = 2 (92)				
ALAMET01	AM	DL/LD		**S4**
AQAROZ	AI		hydrate	VA-class
AQARUF	IA			VA-class
AQASAM	IV		0.21 hydrate	VA-class
AQASEQ	VI		0.22 hydrate	VA-class
AQASIU	VV		hydrate	VA-class
BAPBEZ10	PM		hydrate	**T5**
BELCUQ	EE			**T5**
BIBVOX	PV		hydrate	**T5**

BOFZOL	LE			**S4**
BUHGIU	PG		hydrate	**T5**
BURLIJ	AD			**S5**
BURLOP	ED			**S4**
CAQTOD	MA		2-propanol solvate	**S4**, *Z*' = 2
CAZGOA01	VS			VA-class, *Z*' = 3
CIJGUX	VE			**S5**
COCGEG	AF		2 (2-propanol) solvate	**S5**
COCGIK	VF		2-propanol solvate	**S4**, *Z*' = 4
COCGOQ	LF		0.5 hydrate 2-propanol solvate	**S4**, *Z*' = 2
DABQOM	nGfG		1,2-dimethoxybenzene clathrate	**S4**
DEZQOO	AL		0.5 hydrate	**T5**
DIYZIU	RE		hydrate	anti
DUHKEW	ED		hydrate	**T4**
ETITUW	IL		0.91 hydrate	**T5** tubular, *Z*' = 2
EYIVAJ	SV			**T5**
FABYEM10	AnV	DL/LD		**S4**
FEHPAK	LA		benzyl methyl sulfoxide clathrate	**S4**
FEHPEO	LA		isobutyl methyl sulfoxide clathrate	**S4**
FOBLUE	VS		trihydrate	**T4**
FOBXAW	NV		1.33 hydrate	no layer, *Z*' = 3
FUJZUF	AW		hydrate	unique[f]
FULGEY02	WG		hydrate	**T5** tubular
GEHTAP	FW		0.75 hydrate	**T5** tubular, *Z*' = 4
GLUGLY	EG			**S5**
GLYGLY01	GG			**T4**m (β-polymorph)
GUFQON06	GS			no layer
GUKVUD	SL			**T5**
GUYTAV	nGfG		1,2-dimethoxyethane hydrate clathrate	anti
HIQWAF	LL		2-methyl-1-propanol solvate	**S5**, *Z*' = 2
HIZCOJ	LI		2.5 hydrate	no layer
HUZVON	fGfG	*SR*	dimethylformamide clathrate	**S4**
HUZVUT	fGfG	*SR*	acetamide clathrate	**S4**
IDUZOW	LL		0.87 hydrate	**T5** tubular, *Z*' = 2
IDUZUC	LF		0.86 hydrate	**T5** tubular, *Z*' = 2
IFABAS	FL		1.26 hydrate	**T5** tubular, *Z*' = 2
IFABEW	FF		2.47 hydrate	**T5** tubular
JENTOL01	fGfG		(*S*)-isopropylphenylsulfoxide clathrate	**S4**
JUCSEF01	LV		2-propanol solvate.	**S5**
JUKMEH	LY			**S5**
JUKMOR	HL			**T5**
JUQQIV	LL		1-propanol 2-propanol solvate	anti, *Z*' = 2
KIXBID	SN		hydrate	**T4**
KIYHOP	AS			**S4**m
MAVVUA	nGfG		-[g]	anti, *Z*' = 2
MAZXUH	IS		0.33 hydrate	**T4**
MAZYES	MS		0.34 hydrate	**T4**, *Z*' = 2
METMET	MM			**S4**
NAFZID	LV		0.75 hydrate	no layer, *Z*' = 4
NAYJOM	HG		0.5 hydrate	no layer, *Z*' = 2
NAYZET	VA		0.33 acetonitrile solvate 0.29 hydrate	VA-class, *Z*' = 3
NOTSIY	fGfG		0.5 (*R*,*R*)-bis((o-methylsulfinyl)benzyl)ether clathr.	**S4**
OLOGEB	MA			VA-class, *Z*' = 7
PAJFIQ	SF			**T5**
PAJPUM	FI		0.88 hydrate	**S5**/**T5** hybrid, *Z*' = 2
PRSARH	PSar		hydrate	anti
RAVZAQ	HM			**T5**
RAVZOE	**HD**		trihydrate	**T4**
RAWBAT	AH		ethanol solvate 0.5 hydrate	anti
SAMWOT	fGfG		(4-fluorobenzyl)methylsulfoxide clathrate	**S5**
SAMWUZ	fGfG		(4-bromobenzyl)methylsulfoxide clathrate	**S5**

SBLCYS	bCbC			**S5**
SEGYOS	RV+		acetate	**S5**
SEYWAU	KV+		chloride	**S5**
SEYWEY	PI		hydrate	**T5**
SOJPAI	PY		hydrate	**T5**
SUWLIF	LV		methanol solvate	**S5**
SUWLOL	LV		ethanol solvate	**S5**, *Z*' = 4
TARKUT	MN			**S5**
TEKNAY	HA		dihydrate	**S5**
TELVOV01	LA		dimethylsulfoxide clathrate	**S4**
TIPTOB	VQ			**S5**
TIPTUH	EV			**T5**
VIFFEW	FY			**S5**
WIRYEB	VA			VA-class
XEGNAY	FV			**S5**
XEJQOS	fGfG		rac-benzyl methyl sulfoxide clathrate	**S4**, *Z*' = 2
XUDVOH	AV			VA-class
XUDWAU	AV		0.25 2-propanol solvate 0.22 hydrate	VA-class, *Z*'= 4
YAMHUP	RS+		acetate hydrate	**S4**
YICGUM	KL+		acetate 0.5 acetic acid solvate 0.5 hydrate	**S5**, *Z*' = 2
YORPEA	LL		dimethyl sulfoxide solvate	**S5**
ZILDON	GQ		hydrate	**S5**
N = 3 (25)				
ALAALA	AA			**T5** tubular
ALAGLY	AG			**S5**
EMIPAR	AM		0.5 hydrate	**S4/T5** hybrid, *Z*' = 2
ETIWIN	LI		0.75 hydrate	no layer, *Z*' = 4
EWOVAN	AT			**S5**m
GLDLPA	GF	L/D		anti
GLHPRA	GhP			no layer
GLTHRE	GT	L/D	hydrate	**S5**
GLYALB	GA			no layer
GLYDLA	GA	L/D		no layer
GLYGLY04	GG			anti (α-polymorph)
GLYLEU10	GL			**T5**
JAXBUG	LS			**T5**m, tubular
JAYFOF	AB			**T5/T5/T4** hybrid, *Z*' = 3
LALLSE	AS			**S4**
MAPKOE	TA			**T4** tubular
MAZYAO	FS	LD		**T5**m
QQQEVJ01	GF			**T5**
SEHGES	GP		0.5 hydrate	unique, *Z*' = 2
SERGLY	SG			**T4**m
WEVWOK	GV			**T5**, *Z*' = 7
WILTIV	fGfG	*SR*		**T5**m
XEGHOG	GL	L/D		unique
XOSHOC	BA			**T5** tubular
XOSHUI	AB		0.33 hydrate	**T5/T5/T4** hybrid, *Z*' = 3

[a] Refcode in the CSD (Allen, 2002), refcodes in bold constituted the database used by Suresh & Vijayan (1985).
[b] Peptide sequence given with one-letter amino acid abbreviations. Uncommon acids: nG = (1-naphtyl)glycine; fG = phenylglycine; nV = norvaline; B = 2-amino butyric acid; hP = hydroxyproline; bC = *S*-benzylcysteine; Sar = sarcosine. An additional '+', as in RS+, indicates a peptide net charge of +1, bold type face indicates a double zwitterion.
[c] Stereochemistry given when deviating from LL, DD, L or D.
[d] The basic pattern is given when present. A succeeding 'm', as in **T5**m, indicates a modified pattern. One or two succeeding stars indicates one or two missing *C*(8) chains, **T5**-**S5**** (for *N* = 0) means that either classification could be justified, parentheses surround tentative classifications.
[f] unique = parallel H bond pattern occurring in this structure only.
[g] 0.5 (*S*)-methyl 2-hydroxy-3,3-dimethylbutyrate methanol clathrate 0.5 hydrate

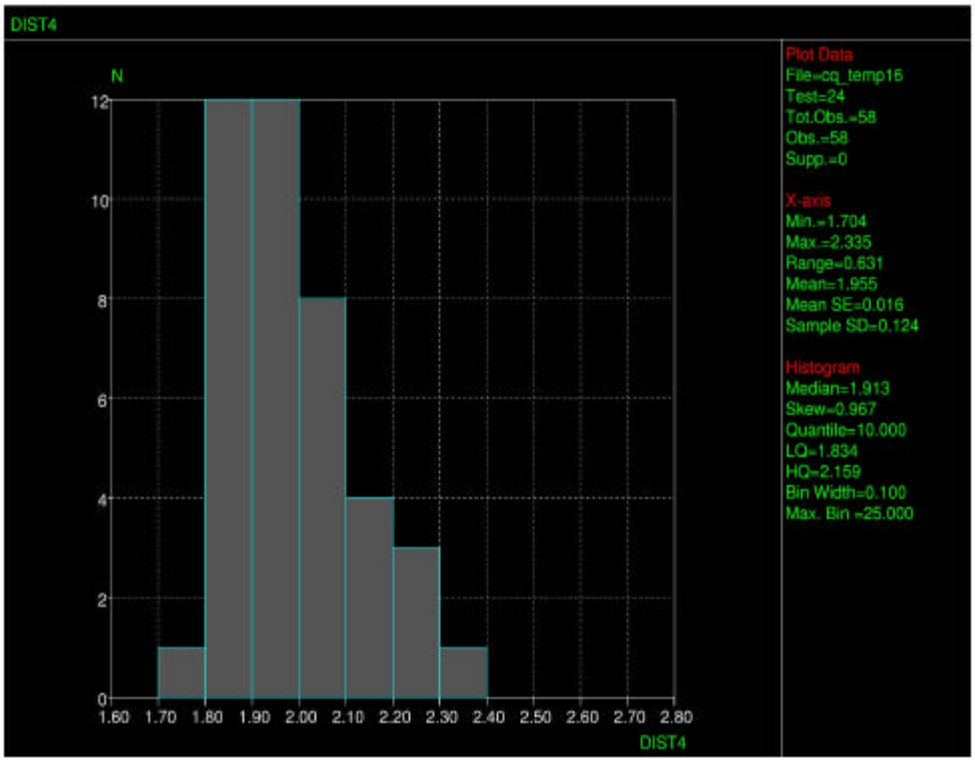


Figure 1S

Distribution of N-H···O distances for *C*(5) chains in **T5** and **S5** structures. N-H distances were normalized to 1.009 Å.

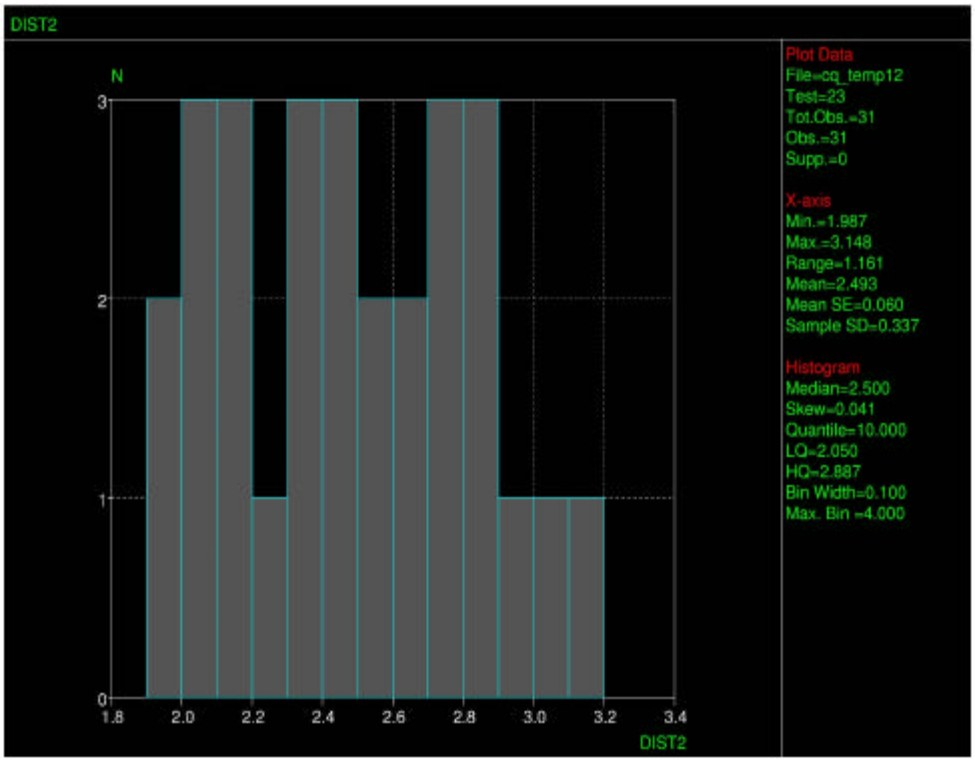


Figure 2S

Distribution of N-H···O distances for *C*(4) chains in **T4** and **S4** structures. N-H distances were normalized to 1.009 Å.

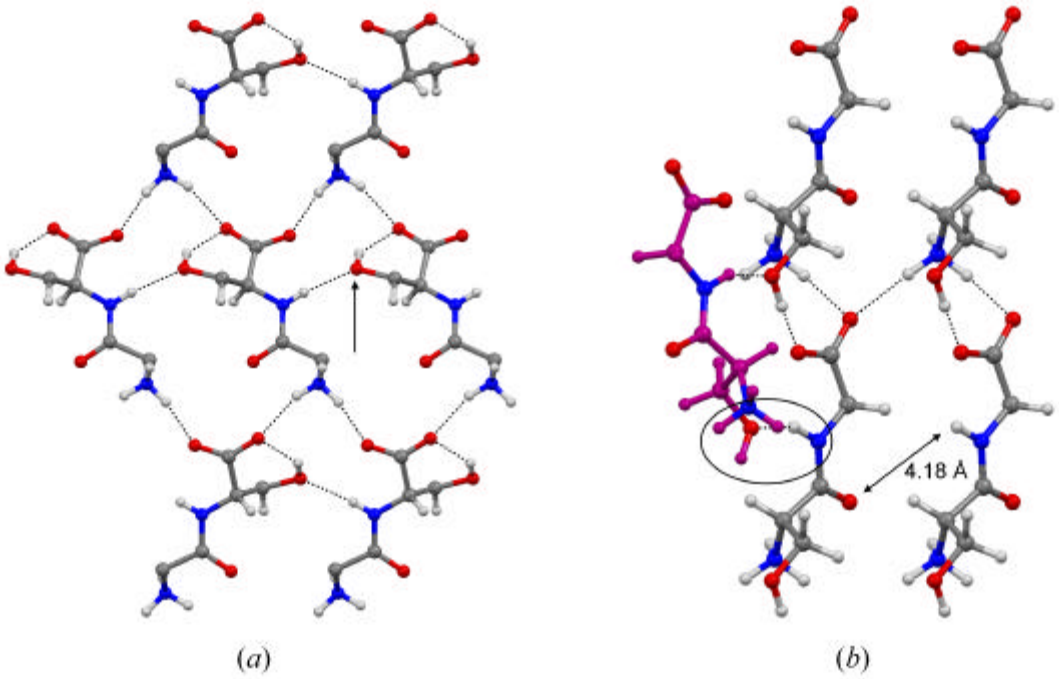
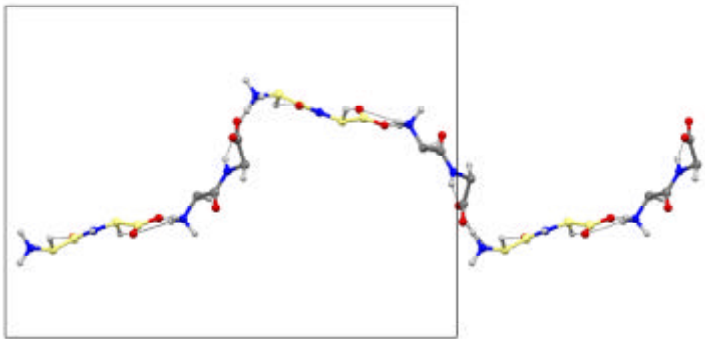
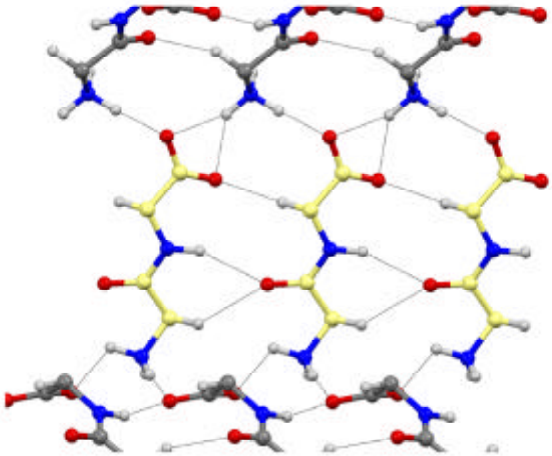


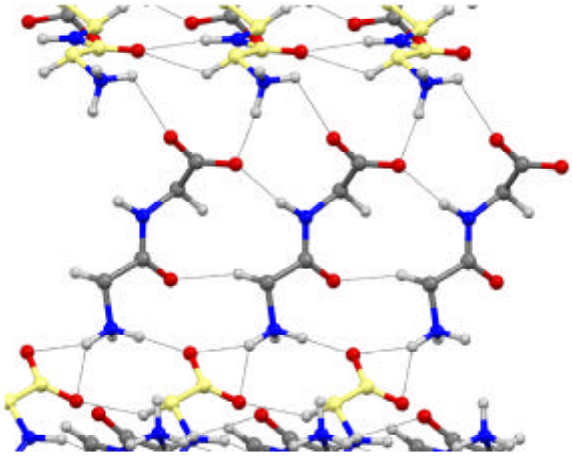
Figure 3S
(a) Example of modified **S5** pattern in the crystal structure of Ala-Thr (Netland *et al*., 2004). The side-chain hydroxyl group (arrow) has inserted itself into what would normally have been a direct bond between the amide >N-H and the carboxylate group. Side chain methyl groups have been omitted for clarity. (b) Modified **T4** pattern in the crystal structure of Ser-Gly (Jones *et al*., 1978a). The peptide is twisted so that the indicated >N-H···O=C< distance becomes too long for a hydrogen bond. Instead, the amide >N-H is donated to a side-chain hydroxyl group (circled) in a peptide molecule in the adjacent layer (C and H atoms coloured in magenta).



(*a*)



(*b*)



(*c*)

Figure 4S

The structure of Ala-Met hemihydrate (Görbitz, 2003a), a rare example of a hybrid structure. (a) A wave-like main chain "layer" viewed along the crystallographic *a*-axis. Side chains and water molecules have been removed; main-chain C-atoms in the two molecules in the asymmetric unit have different colours. (b) Hydrogen bonding of molecule 1 showing typical connectivity of the **S4** pattern. (c) Hydrogen bonding of molecule 2 showing typical connectivity of the **T5** pattern.

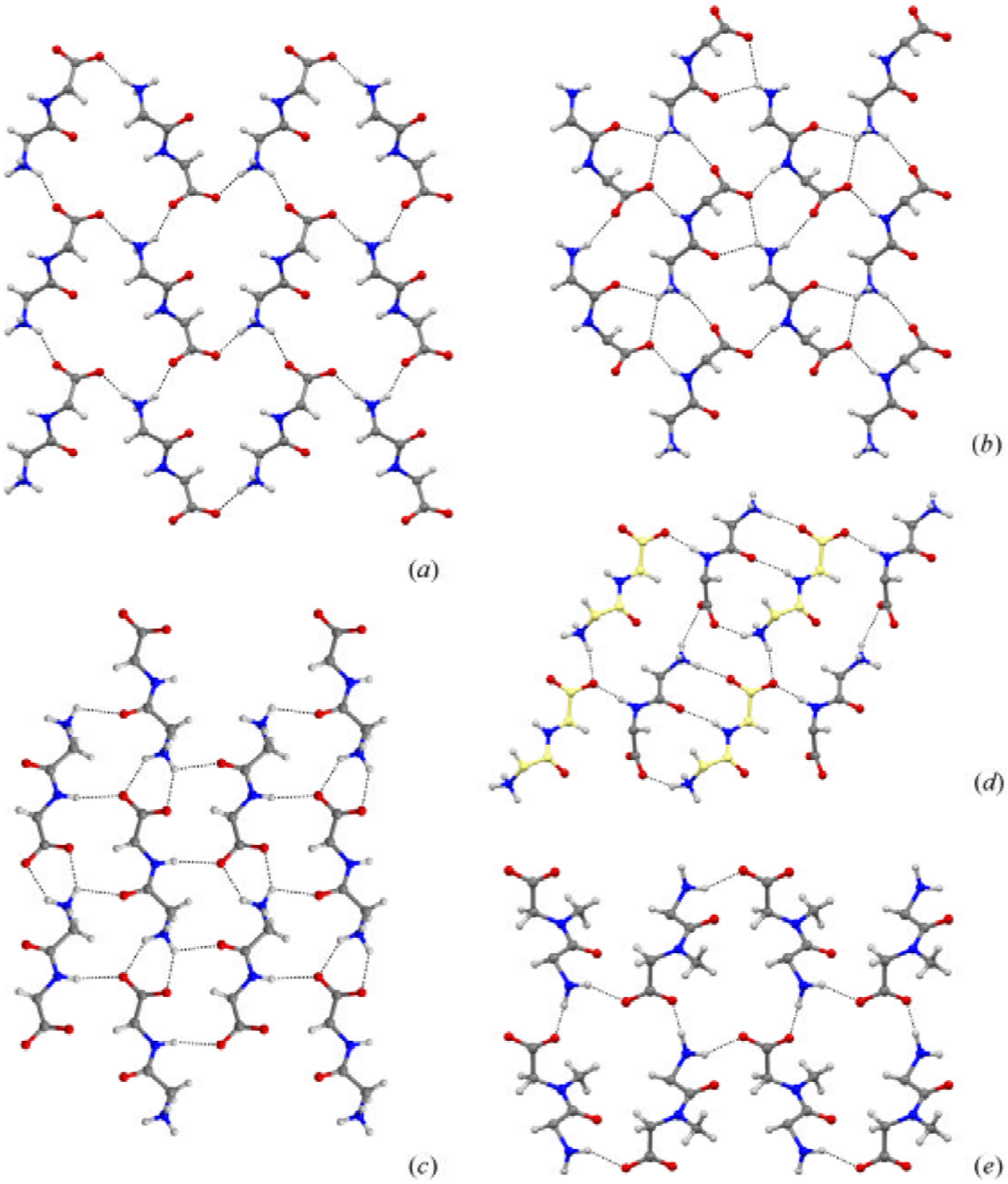
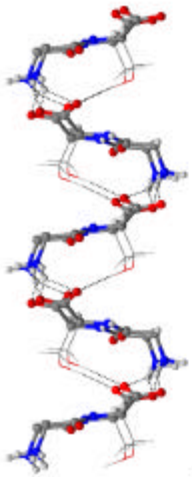
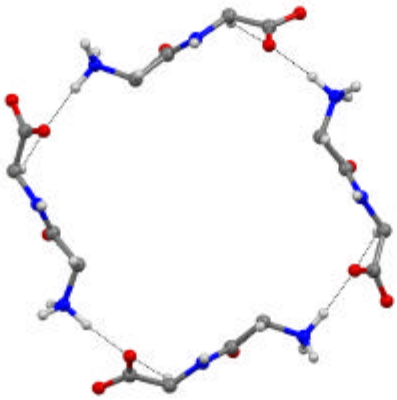


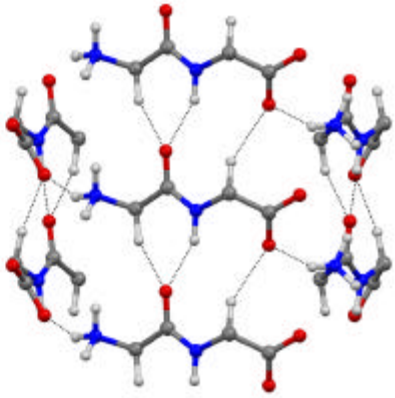
Figure 5S
Antiparallel hydrogen bonding patterns in
(a) (*R*)-nGly-(*R*)-fGly 1,2-dimethoxyethane monohydrate clathrate (Akazome *et al*., 2002), (b) Arg-Glu hydrate (Eggleston & Hodgson, 1985), (c) Gly-DL-Phe (Marsh *et al*., 1976), (d) Leu-Leu ethanol solvate (Görbitz, 1998) with *Z*' = 2, (e) Pro-Sar hydrate (Kojima *et al*., 1980).

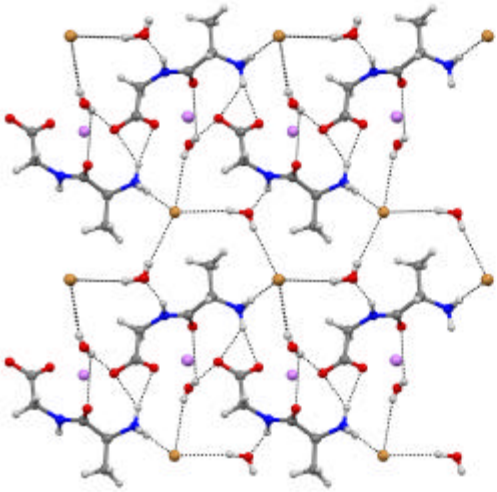


(*a*)

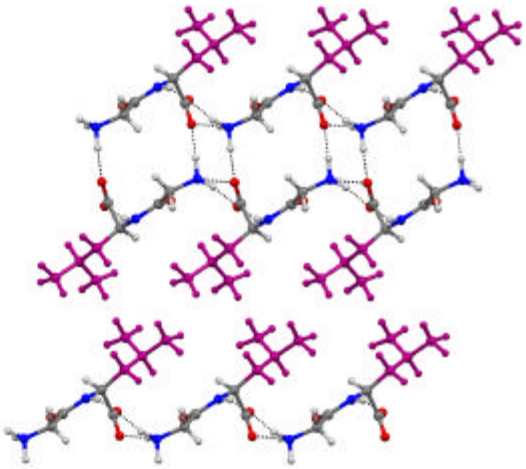


(*b*)

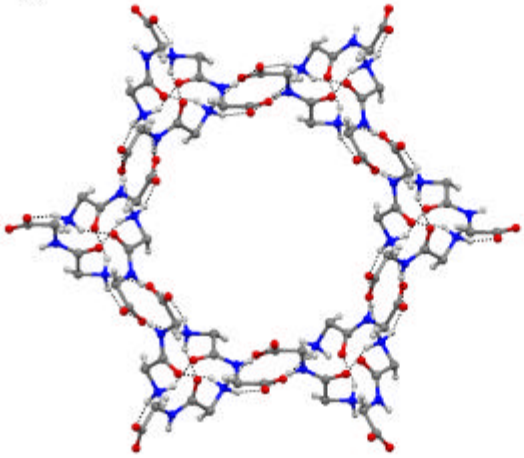


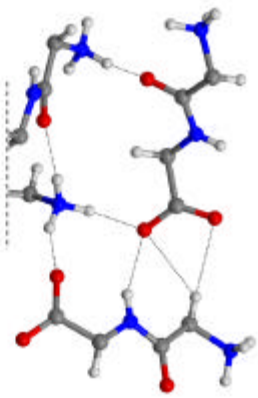


(*c*)



(*d*)





(*e*)

Figure 6S

Selected structures with special features.
(a) Ser-Ala (Jones *et al*., 1978b) with unique, heavily undulated **S4** layers. The Ala side chain has been removed, the Ser side chain is shown in wireframe representation.
(b) Tubular **T4** hydrogen bonding in Thr-Ala (Görbitz, 2005) shown along the tetragonal axis (left) and tilted 90° around the *x*-axis (right). Side chains and the rear side of the tube in (b) have been hidden for clarity.
(c) Ala-Gly·LiBr hydrate (Declercq *et al*., 1971), an unusual N =1 (three-centered interaction) peptide structure incorporating a metal salt.
(d) The **T5** structure of Gly-Leu (Patthabi *et al*., 1974) where the lack of a normal side chain for the first residue renders direct hydrogen-bonding contact between adjacent main-chain layers possible. The structure may be compared with the structure of His-Leu (Krause *et al*., 1993) (Fig. 10 in the main body of the paper).
(e) Ile-Ala (Görbitz, 2003b), a member of the Val-Ala class that represent the largest group of N = 2 structures devoid of layers. The left-hand view is along the hexagonal axis, a detail of the hydrogen bonding pattern showing the characteristic *syn* amide N-H···carboxylate and C-H···carboxylate interactions is given to the right. Side chains have been removed for clarity.

References

Akazome, M., Takahashi, T., Sonobe, R. & Ogura, K. (2002). *Tetrahedron* **58**, 8857-8861.
Allen, F. H. (2002). *Acta Cryst.* B**58**, 380-388.
Declercq, J. P., Meulemans, R., Piret, P. & van Meerssche, M. (1971). *Acta Cryst.* B**27**, 539-544.
Eggleston, D. S. & Hodgson, D. J. (1985). *Int. J. Pept. Protein Res*. **25**, 242-253.
Görbitz, C. H. (1998). *Acta Chem. Scand.* **52**, 1343-1349.
Görbitz, C. H. (2003a). *Acta Cryst.* C**59**, o730-o732.
Görbitz, C. H. (2003b). *New J. Chem.* **27**, 1789-1793.
Görbitz, C. H. (2005). *Acta Cryst.* E**61**, o2012-o2014.
Jones, P. G., Falvello, L. & Kennard, O. (1978a). *Acta Cryst.* B**34**, 2379-2381.
Jones, P. G., Falvello, L. & Kennard, O. (1978b). *Acta Cryst.* B**34**, 1939-1942.
Kojima, T., Kido, T., Itoh, H., Yamane, T. & Ashida, T. (1980). *Acta Cryst.* C**36**, 326-331.
Krause, J. A., Baures, P. W. & Eggleston, D. S. (1993). *Acta Cryst.* B**49**, 123-130.
Marsh, R. E., Ramakumar, S. & Venkatesan, K. (1976). *Acta Cryst.* B**32**, 66-70.
Netland, K. A., Andresen, K., Görbitz, C. H. & Dalhus, B. (2004). *Acta Cryst.* E**60**, o951-o953.
Patthabi, V., Venkatesan, K. & Hall, S. R. (1974). *J. Chem. Soc., Perkin Trans. 2*, 1722-1727.
Suresh, C. G. & Vijayan, M. (1985). *Int. J. Peptide Protein Res*. **26**, 311-328.