



STRUCTURAL SCIENCE
CRYSTAL ENGINEERING
MATERIALS

Volume 74 (2018)

Supporting information for article:

Hydrophobic dipeptides: the final piece in the puzzle

Carl Henrik Görbitz

Hydrophobic dipeptides: the final piece in the puzzle

Carl Henrik Görbitz

Department of Chemistry, University of Oslo, Norway

Supporting information

	Page
Figure S1 Standard and non-standard S4 hydrogen bonding pattern	2
Figure S2 Selection of <i>anti</i> hydrogen bonding patterns	3
Figure S3 S5/T5 hybrid hydrogen bonding pattern in the crystal structure of Phe-Ile 0.88 hydrate	4
Table S1 Hydrogen bonding patterns in the crystal structures of hydrophobic dipeptides (short version)	
Table S2 Hydrogen bonding patterns in the crystal structures of hydrophobic dipeptides (extended version including redeterminations and pseudopolymorphs)	6
Table S3 Hydrogen bonding patterns in the crystal structures of all dipeptides	11
Table S4 Torsion angles in dipeptide structures for the main structural classes	15
References	18

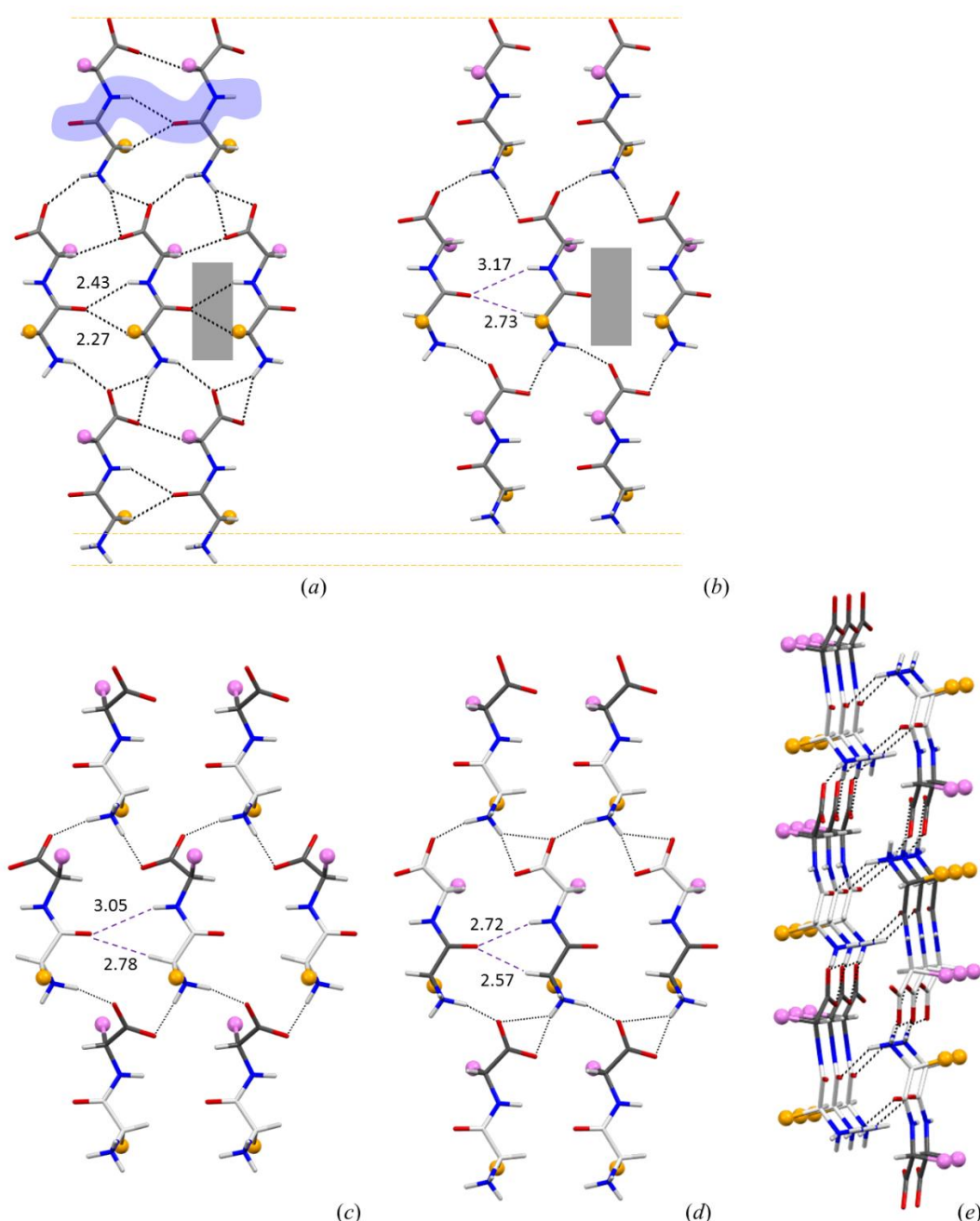


Figure S1

(a) Standard **S4** hydrogen bonding pattern in the structure of Phe-Phe methanol solvate (CSD refcode JOQLIM)(Mason *et al.*, 2014) with a *C*(4) chain involving the amide >N-H donor (blue shade). There is also a *C*(4) chain involving the neighbouring C^α-H donor. (b) Due to the presence of a large co-crystallized molecule in D-fGly-D-fGly (*R*)-2-chlorophenyl methyl sulfoxide clathrate (XEJQIM, in the illustration inverted to the L-L enantiomer)(Akazome *et al.*, 2000), head-to tail chains are more widely separated (grey box) so that H...O distances (given in Å after normalization of N-H covalent bonds to 0.88 Å and C-H bonds to 1.00 Å) become long for hydrogen bonds. The relationship with the structure in (a) is nevertheless obvious, so this is still referred to as a **S4** pattern. Expansion along the horizontal axis leads to contraction along the vertical axis, as indicated by the dashed, orange lines. (c) Hydrogen bonding in (*S*)-fGly-(*R*)-fGly formamide solvate (Akazome *et al.*, 2002). C-atoms of L- and D-amino acids are coloured in light and dark grey, respectively. The peptide conformation is rather folded, and the sheet is isolated in the crystal structure. (d) **S4** hydrogen bonding in the racemate LD-Ala-DL-Met (Murali *et al.*, 1986), chirality colour code as in (c). Two such sheets, related by a center-of-symmetry, make a hydrophilic layer in the structure, as seen edge-on in (e).

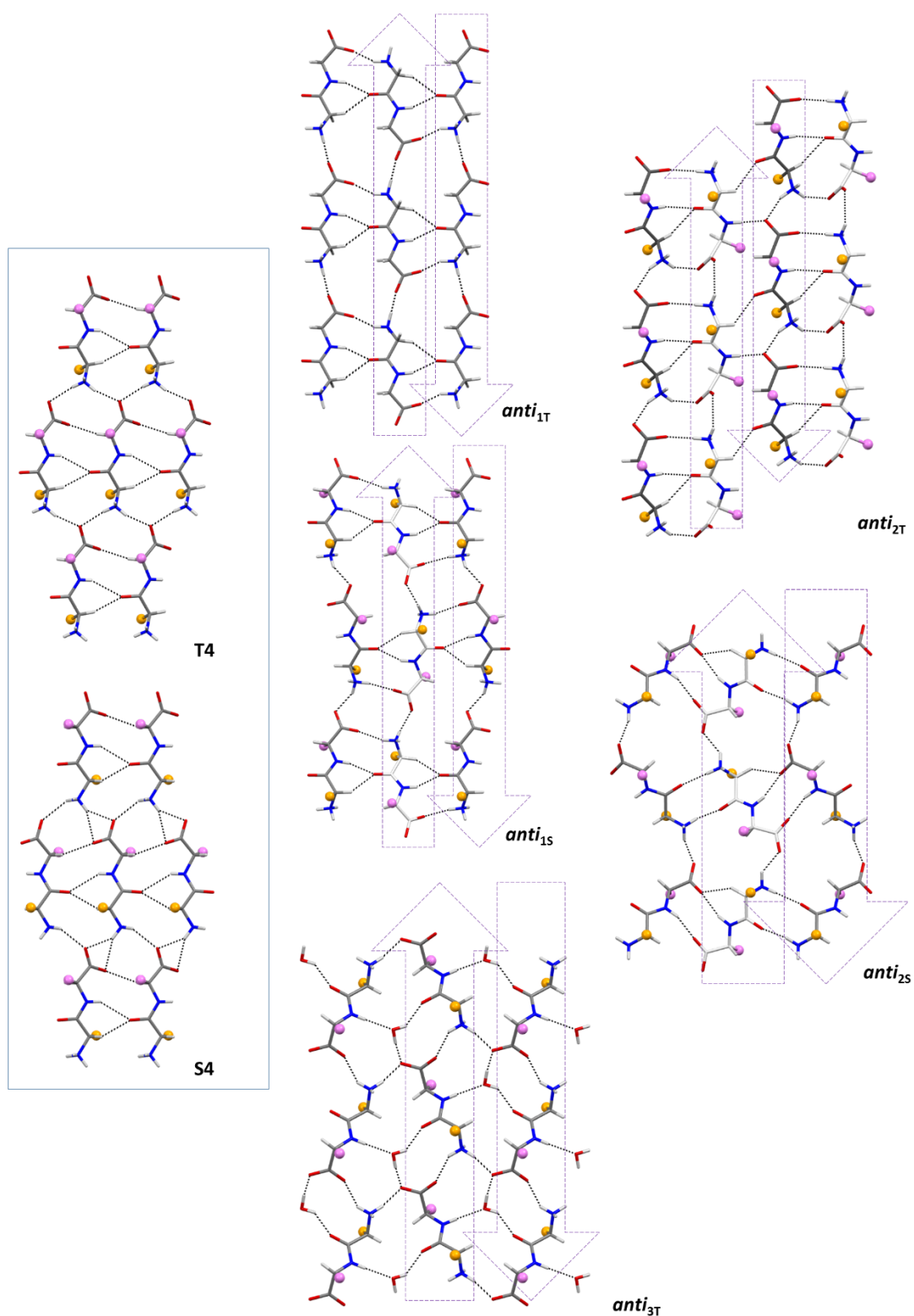


Figure S2

Selection of *anti* hydrogen bonding patterns. The subscript gives a running number and relationship to one of the standard patterns **T4** or **S4** (inside box). The *anti*_{1T} (observed only for α -Gly-Gly)(Kvick *et al.*, 1977, Biswas *et al.*, 1968) and *anti*_{1S} patterns can be derived from the standard patterns by flipping every second head-to-tail chain in the opposite direction, as indicated by the large, open arrows. In doing so only one C(8) chain is retained. Like *anti*_{1S}, the *anti*_{2T} and *anti*_{2S} patterns have two molecules in the asymmetric unit (drawn with light and dark grey C-atoms), while *anti*_{3T} has an integrated water molecule.

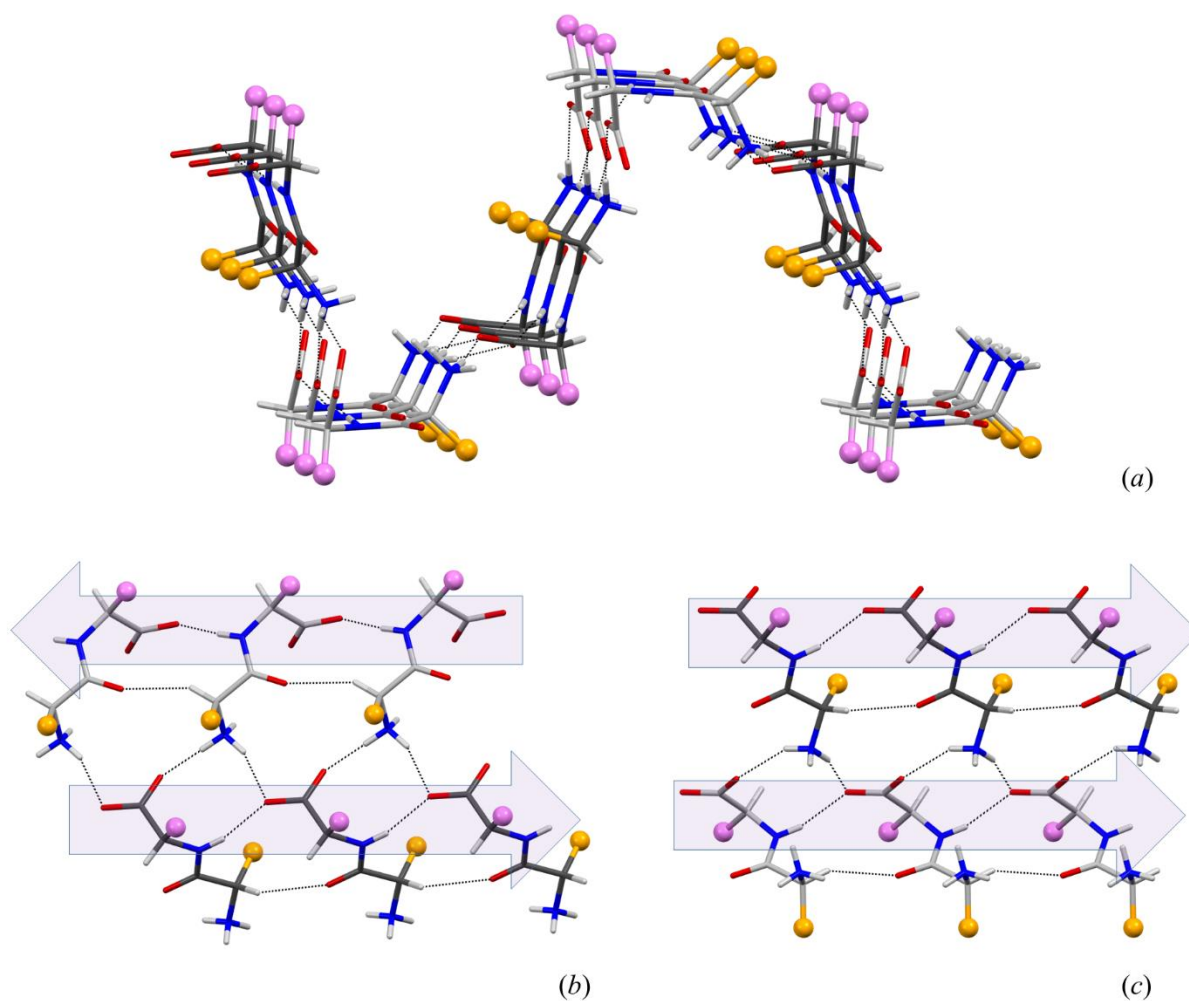


Figure S3

S5/T5 hybrid hydrogen bonding pattern in the crystal structure of Phe-Ile 0.88 hydrate (PAJPUM)(Görlitz, 2004b). (a) Hydrogen bonds between the main chains of the two molecules in the asymmetric unit (light and dark grey C-atoms, respectively) generate a strongly undulating 2-dimensional sheet. (b) One interface between rows of molecules in C(5) chains showing characteristics of a **S5** pattern in that neighbouring chains (large, open arrows) are antiparallel. (c) Second interface showing characteristics of a **T5** pattern with parallel C(5) chains.

Table S1Hydrogen bonding patterns in the crystal structures of hydrophobic dipeptides.^a

Sequence ^b	Refcode ^c	Comment	Packing ^d	Sp. gr. ^e	Z'	T (K)	Year
<i>Peptides with combinations of Ala, Val, Ile, Leu and Phe</i>							
Ala-Ala	ALAALA		T5-t	I4	1	295	1971
Ala-Val	XUDVOH		VA	P6 ₁	1	150	2002
	XUDWAU	0.25 2-propanol 0.22 H ₂ O	VA	P6 ₁	4	150	2002
Ala-Ile	AQARQZ	0.47 H ₂ O	VA	P6 ₁	1	150	2003
Ala-Leu	DEZQOO	0.50 H ₂ O	T5	C2	1	150	1999
Ala-Phe	COCGEG	2.00 2-propanol, A mode ^f	S5	P2 ₁ 2 ₁ 2 ₁	1	150	1999
Val-Ala	WIRYEB		VA	P6 ₁	1	120	1996
	NAYZET	0.35 acetonitrile 0.29 H ₂ O	VA	P2 ₁	3	105	2005
Val-Val	AQASIU	1.03 H ₂ O	VA	P6 ₁	1	150	2003
Val-Ile	AQASEQ	0.22 H ₂ O	VA	P6 ₁	1	105	2003
Val-Leu	(I) this work	1.00 acetonitrile, A mode	S5	P2 ₁ 2 ₁ 2 ₁	1	105	2018
Val-Phe	COCGIK	1.00 2-propanol	S4	P2 ₁	4	150	1999
	MOBYAD	2.00 H ₂ O, similar to Nva-Phe (VIKWUJ)	T5**	P2 ₁ 2 ₁ 2 ₁	1	150	2002
	MOBYEH	3.00 H ₂ O	O	P2 ₁	8	150	2002
Ile-Ala	AQARUF		VA	P6 ₁	1	105	2003
Ile-Val	AQASAM	0.21 H ₂ O	VA	P6 ₁	1	105	2003
Ile-Ile	YAGZOW	2.00 H ₂ O	(S5*)	P2 ₁ 2 ₁ 2 ₁	1	105	2004
Ile-Leu	ETITUW	0.91 H ₂ O	T5-t	C2	2	105	2004
Ile-Phe	ETONIK	2.00 H ₂ O	T5**	P2 ₁	1	105	2004
Leu-Ala	TELVOV	1.00 DMSO	S4	P2 ₁ 2 ₁ 2 ₁	1	295	1996
	FEHPAK	1.00 benzyl methyl sulfoxide	S4	P2 ₁	1	298	2005
	RAVMOQ	4.00 H ₂ O	S5*	P2 ₁ 2 ₁ 2 ₁	1	150	1997
Leu-Val	NAFZID	0.75 H ₂ O, similar to Leu-Ile (ETIWIN)	O	P6 ₂	4	120	1996
	JUCSEF01	1.00 2-propanol, Ac mode	S5	P2 ₁ 2 ₁ 2 ₁	1	150	1999
	SUWLIF	1.00 methanol	S5	P2 ₁	1	150	1999
	SUWLLO	1.00 ethanol	S5	P2 ₁	4	150	1999
Leu-Ile	ETIWIN	0.75 H ₂ O, similar to Leu-Val (NAFZID)	O	P6 ₂	4	105	2004
	HIZCOJ	2.50 H ₂ O	O	P2 ₁ 2 ₁ 2 ₁	2	110	2008
	IKOMOM	1.00 trifluoroethanol	S5	P2 ₁	2	120	2016
Leu-Leu	YORPEA	1.00 DMSO	S5	P2 ₁	1	295	1994
	HIQWAF	1.00 2-methyl-1-propanol, A mode, Z' = 2	S5	P2 ₁ 2 ₁ 2 ₁	2	150	1999
	IDUZOW	0.87 H ₂ O	T5-t	P2 ₁ 2 ₁ 2 ₁	2	150	2001
	JUQQIV	1.00 ethanol	anti ₂₇	P2 ₁	2	150	1998
Leu-Phe	COCGOQ	1.00 2-propanol 0.10 H ₂ O	S4	P2 ₁	2	150	1999
	IDUZUC	0.86 H ₂ O	T5-t	P2 ₁	2	150	2001
Phe-Ala	QIMBUJ	2.00 H ₂ O	screw	P2 ₁ 2 ₁ 2 ₁	1	150	2001
Phe-Val	XEGNAY		S5	P2 ₁ 2 ₁ 2 ₁	1	150	2000
Phe-Ile	PAJPUM	0.88 H ₂ O	S5/T5	P2 ₁	2	105	2004
Phe-Leu	IFABAS	1.26 H ₂ O	T5-t	P2 ₁ 2 ₁ 2 ₁	2	150	2001
Phe-Phe	IFABEW	2.47 H ₂ O	T5-t	P6 ₁	1	150	2001
	JOQLIM	2.00 methanol	S4	P2 ₁ 2 ₁ 2 ₁	1	298	2014
<i>Other hydrophobic dipeptides with S5 sheets</i>							
Phe*-Phe*	ITOBFA		S5	P2 ₁	3	100	2016
Ile-fGly	OKOZII	1.00 methyl o-tolyl sulfoxide, A mode	S5	P2 ₁ 2 ₁ 2 ₁	1	173	2010
fGly-fGly	SAMWUZ	1.0 (4-bromobenzyl)methylsulfoxide	S5	P2 ₁	1	173	2004
	SAMWOT	1.0 (4-fluorobenzyl)methylsulfoxide, A mode	S5	P2 ₁ 2 ₁ 2 ₁	1	173	2004
bCys-bCys	SBLCYS	Bc mode	S5	P2 ₁ 2 ₁ 2 ₁	1	295	1975

^a Excluding redeterminations and pseudopolymorphs. ^b Phe* = Phe with -CH₂- replaced by -C₂H₄-, fGly = phenylglycine; bCys = S-benzylcysteine. ^c Rows in grey shade with refcodes in bold share S5 pattern and P2₁2₁2₁ space group with (I) presented here. ^d Designators from Figure 1 of the main body of the paper; yellow, blue and red fill colour indicates structural family as in Fig. 6(a). VA is the nanotubular Val-Ala class, while O means absence of a clear pattern. The S5* pattern of Ile-Ile dehydrate (YAGZOW)(Görbitz, 2004a) is only tentative and has been put in parenthesis. For Phe-Ala (QIMBUJ)(Görbitz, 2001) "screw" means presence of a single, isolated C(8) chain of molecules related by a screw axis. ^e Space group. ^f The various modes are described in (Görbitz, 2010).

Table S2

Hydrogen bonding patterns in the crystal structures of hydrophobic dipeptides [1]

Sequence	Refcode	Comment	Packing	Sp. gr.	Z'	T (K)	Year
Peptides with combinations of Ala, Val, Ile, Leu and Phe							
Ala-Ala	ALAALA		T5-t	I4	1	295	1971
	ALAALA01		T5-t	I4	1	293	2010
	TUWWEO	0.002 O2	T5-t	I4	1	293	2010
Ala-Val	XUDVIB	0.35 acetonitrile	VA	P61	1	150	2002
	XUDVUN	0.75 methanol	VA	P61	1	150	2002
	XUDVOH		VA	P61	1	150	2002
	XUDVOH01		VA	P61	1	293	2004
	XUDVOH04		VA	P61	1	293	2010
	XUDVOH05	at 0.2 Gpa	VA	P61	1	293	2010
	EPATUM	unknown solvate	VA	P61	1	296	2014
	XUDWAU	0.25 2-propanol 0.22 H2O	VA	P61	4	150	2002
Ala-Ile	AQAR0Z	0.47 H2O	VA	P61	1	150	2003
	HEGLOV		VA	P61	1	293	2006
Ala-Leu	DEZQOO	0.50 H2O	T5	C2	1	150	1999
Ala-Phe	COCGEG	2.00 2-propanol, A mode	S5	P212121	1	150	1999
Val-Ala	WIRYEB		VA	P61	1	120	1996
	WIRYEB02		VA	P61	1	293	2004
	NAYZIX	0.12 acetonitrile	VA	P61	1	105	2005
	NAYZET	0.35 acetonitrile 0.29 H2O	VA	P21	3	105	2005
Val-Val	AQASIU	1.03 H2O	VA	P61	1	150	2003
	HEGLUB		VA	P61	1	293	2006
Val-Ile	AQASEQ	0.22 H2O	VA	P61	1	105	2003
	HEGMEM		VA	P61	1	293	2006
	HEGMEM01		VA	P61	1	293	2010
Val-Leu	<i>This work</i>	1.00 acetonitrile, A mode	S5	P212121	1	105	2018
Val-Phe	COCGIK	1.00 2-propanol	S4	P21	4	150	1999
	MOBYAD	2.00 H2O, similar to Nva-Phe (VIKWUJ)	T5**	P212121	1	150	2002
	MOBYEH	3.00 H2O	O	P21	8	150	2002

Ile-Ala	AQARUF		VA	P61	1	105	2003
	AQARUF01		VA	P61	1	293	2006
Ile-Val	AQASAM	0.21 H2O	VA	P61	1	105	2003
	HEGMAI		VA	P61	1	293	2006
Ile-Ile	YAGZOW	2.00 H2O	(S5*)	P212121	1	105	2004
Ile-Leu	ETITUW	0.91 H2O	T5-t	C2	2	105	2004
Ile-Phe	ETONIK	2.00 H2O	T5**	P21	1	105	2004
Leu-Ala	TELVOV	1.00 DMSO	S4	P212121	1	295	1996
	TELVOV01	1.00 DMSO	S4	P212121	1	298	2005
	FEHPEO	1.00 isobutyl methyl sulfoxide	S4	P212121	1	298	2005
	FEHPAK	1.00 benzyl methyl sulfoxide	S4	P21	1	298	2005
	RAVMOQ	4.00 H2O	S5*	P212121	1	150	1997
Leu-Val	NAFZID	0.75 H2O, similar to Leu-Ile (ETIWIN)	O	P62	4	120	1996
	JUCSEF01	1.00 2-propanol, Ac mode	S5	P212121	1	150	1999
	SUWLIF	1.00 methanol	S5	P21	1	150	1999
	SUWLOL	1.00 ethanol	S5	P21	4	150	1999
Leu-Ile	ETIWIN	0.75 H2O, similar to Leu-Val (NAFZID)	O	P62	4	105	2004
	HIZCOJ	2.50 H2O	O	P212121	2	110	2008
	IKOMOM	1.00 trifluoroethanol	S5	P21	2	120	2016
Leu-Leu	YORPEA	1.00 DMSO	S5	P21	1	295	1994
	HIQWAF	1.00 2-methyl-1-propanol, A mode, Z' = 2	S5	P212121	2	150	1999
	IDUZOW	0.87 H2O	T5-t	P212121	2	150	2001
	JUQQIV	1.00 ethanol	anti- 2T	P21	2	150	1998
	JUQQOB	1.00 1-propanol	anti- 2T	P21	2	150	1998
	JUQQUH	1.00 2-propanol	anti- 2T	P21	2	150	1998
	JUQRAO	0.55 2-propanol 0.45 1-propanol	anti- 2T	P21	2	150	1998
Leu-Phe	COCGOQ	1.00 2-propanol	S4	P21	2	150	1999
	IDUZUC	0.86 H2O	T5-t	P21	2	150	2001
Phe-Ala	QIMBUJ	2.00 H2O	screw	P212121	1	150	2001
Phe-Val	XEGNAY		S5	P212121	1	150	2000
Phe-Ile	PAJPUM	0.88 H2O	S5/T5	P21	2	105	2004
Phe-Leu	IFABAS	1.26 H2O	T5-t	P212121	2	150	2001

Phe-Phe	IFABEW	2.47 H ₂ O	T5-t	P61	1	150	2001
	OREVAK	1.00 H ₂ O (PXRD)	T5-t	P61	1	298	2014
	JOQLEI	1.60 H ₂ O	T5-t	P61	1	298	2014
	JOQLIM	2.00 methanol	S4	P212121	1	298	2014
Peptides with Met							
Ala-Met	EMIPAR	0.5 H ₂ O	S4/T5	P212121	2	105	2003
Met-Ala	CAQTOD	1.0 2-propanol	S4	P21	2	150	2000
Met-Ala	OLOGEB		VA	P61	7	105	2003
Met-Met	METMET		S4	P212121	1	295	1975
Peptides with Trp							
Ala-Trp	FUJZUF	1.0 H ₂ O	O	P212121	1	295	2000
Phe-Trp	GEHTAP	0.75 H ₂ O	T5-t	P212121	4	105	2006
Ile-Trp	BEQJAJ	2.0 H ₂ O, similar to Val-Phe (MOBYAD)	T5**	P212121	1	193	2004
Trp-Val	BEQJEN	2.0 H ₂ O	S5**	P212121	1	193	2004
VA-class peptides with Abu and/or Nva							
Nva-Val	GAGFIG		VA	P61	1	105	2015
Nva-Leu	GAGFOM		VA	P61	1	105	2015
Nva-Ile	GAGFUS		VA	P61	1	105	2015
Abu-Abu	GAGPAI		VA	P61	1	105	2015
Abu-Nva	GAGPEM		VA	P61	1	105	2015
Abu-Val	GAGPIQ		VA	P61	1	105	2015
Abu-Leu	GAGPOW		VA	P61	1	105	2015
Abu-Ile	GAGPUC		VA	P61	1	105	2015
Nva-Abu	GAGQAJ		VA	P61	1	105	2015
Nva-Nva	GAGQEN		VA	P61	1	105	2015
Other peptides with Abu or Nva							
Ala-Abu	JAYFOF		T5/T5/T4	P21	3	105	2005
Ala-Abu	XOSHUI	0.33 H ₂ O	T5/T5/T4	P21	3	293	2002
Abu-Ala	XOSHOC		T5-t	I4	1	293	2002
Nva-Phe	VIKWUJ	2.00 H ₂ O, similar to Val-Phe (MOBYAD)	T5**	P212121	1	105	2013

Peptides with Phe derivatives							
Phe*-Phe*	ITOEBA		S5	P21	3	100	2016
Phe*-Phe*	ITOBIE	0.5 trifluoroacetate	<i>anti</i>	P212121	4	100	2016
Phe*-Phe*	ITOBOK	0.25 trifluoroacetate 0.25 H3O+ 0.125 H2O	<i>anti</i>	P21	8	100	2016
Peptides with fGly and/or nGly							
Ile-fGly	OKOZQA	1.00 methyl phenyl sulfoxide	S4	P212121	1	173	2010
Ile-fGly	OKOZIY	1.00 methyl o-tolyl sulfoxide, A mode, layer slide	S5	P212121	1	173	2010
Leu-fGly	OKOZEU	1.00 methyl 3,4-dimethylphenyl sulfoxide	S4	P212121	1	173	2010
nGly-fGly	MAVTEI	1.0 (S)-ethyl lactate <i>RR</i>	<i>anti</i> - 2S	P21	2	295	1999
	MAVTIM	0.5 pantolactone 1.0 methanol 0.5 H2O <i>RR</i>	<i>anti</i> - 2S	P21	2	173	1999
	MAVVOU	1.0 (S)-methyl 2-hydroxy-3-methylbutyrate <i>RR</i>	<i>anti</i> - 2S	P21	2	295	1999
	RUGHOQ	1.0 2.0 methanol 1.0 H2O <i>RR</i> conformationally distinct from MAVTEI	<i>anti</i> - 2S	P21	2	173	1996
	DABQOM	1.0 1,2-dimethoxybenzene <i>RR</i>	S4	P212121	1	295	1997
	RUGHIK	1.0 2.0 methanol 1.0 H2O <i>RR</i> layer slide compared to DABQOM	S4	P212121	1	173	1996
	GUYSUO	1.0 1,2 diethyl ether 1.0 H2O <i>RR</i>	<i>anti</i> - 3T	P212121	1	298	2002
	GUYTAV	1.0 1,2-dimethoxyethane 1.0 H2O <i>RR</i>	<i>anti</i> - 3T	P212121	1	298	2002
	GUYTEZ	1.0 1-methoxy-2-methylthioethane 1.0 H2O <i>RR</i>	<i>anti</i> - 3T	P212121	1	298	2002
	GUYTID	1.0 1-methoxyethane 1.0 H2O <i>RR</i>	<i>anti</i> - 3T	P212121	1	298	2002
	GUYTOJ	1.0 1,2-dimethoxypropane 1.0 H2O <i>RR</i>	<i>anti</i> - 3T	P212121	1	298	2002
	GUYTUP	1.0 1,3-dimethoxypropane 1.0 H2O <i>RR</i>	<i>anti</i> - 3T	P212121	1	298	2002
	MAVTAE	1.0 (S)-methyl lactate <i>RR</i>	<i>anti</i> - 2S	P212121	2	295	1999
	MAVTOS	0.5 methyl 2-hydroxyvalerate 1.0 methanol <i>RR</i>	<i>anti</i> - 2S	P212121	2	173	1999
	MAVTUY	0.5 methyl (S)-2-hydroxyhexanoate 1.0 methanol <i>RR</i>	<i>anti</i> - 2S	P212121	2	173	1999
	MAVVAG	0.5 (S)-methyl 2-hydroxy-4-methylvalerate 1.0 methanol <i>RR</i>	<i>anti</i> - 2S	P212121	2	295	1999
	MAVVEK	0.5 (S)-methyl mandelate 1.0 H2O <i>RR</i>	<i>anti</i> - 2S	P212121	2	295	1999
	MAVVIO	0.5 methyl 2-hydroxy-3-methylbutyrate 1.0 methanol 0.5 H2O <i>RR</i>	<i>anti</i> - 1S	P212121	2	173	1999
	MAVVUA	0.5 (S)-methyl 2-hydroxy-3,3-dimethylbutyrate 1.0 methanol 0.5 H2O <i>RR</i>	<i>anti</i> - 1S	P212121	2	173	1999
fGly-fGly	XEQQOS	1.0 rac-benzyl methyl sulfoxide <i>RR</i>	S4	P1	2	295	2000
	SAMWUZ	1.0 (4-bromobenzyl)methylsulfoxide <i>RR</i>	S5	P21	1	173	2004
	SAMXAG	1.0 (4-iodobenzyl)methylsulfoxide <i>RR</i>	S5	P21	1	298	2004
	SAMWOT	1.0 (4-fluorobenzyl)methylsulfoxide <i>RR</i> , A mode	S5	P212121	1	173	2004
	XEQQEI	1.0 (S)-methyl 3-tolyl sulfoxide <i>RR</i>	S4	P212121	1	295	2000
	XEQQIM	1.0 (R)-2-chlorophenyl methyl sulfoxide <i>RR</i>	S4	P212121	1	295	2000

JENTOL01	1.00 (S)-isopropylphenylsulfoxide	RR	S4	P212121	1	295	1997
NOTSEU	1.00 (S)-ethyl phenyl sulfoxide	RR	S4	P212121	1	295	1997
NOTSIY	1.00 (S)-ethyl phenyl sulfoxide	RR layer slide compared to NOTSEU	S4	P21212	1	295	1997
WILTIV	SR		T5	P21	1	298	2007
WILTOB	1.00 diethylsulfoxide	SR	S4	P21	1	298	2007
HUZVON	1.00 dimethylformamide	SR	S4	P212121	1	173	2002
HUZVUT	1.00 acetamide clathrate	SR	S4	P212121	1	298	2002
HUZWAA	1.00 N,N-dimethylacetamide	SR	S4	P212121	1	298	2002
WINZEZ	1.00 dimethylsulfoxide	SR	S4	P212121	1	298	2007
WILTUH	1.00 ethyl-methylsulfoxide	SR layer slide from HUZVON/WINZEZ	S4	P212121	1	298	2007
Other hydrophobic dipeptides							
bCys-bCys	SBLCYS	Bc mode	S5	P212121	1	295	1975

[1] Legends as in Table S1 with addition of further codes for *anti* patterns taken from Fig. S2 (the patterns for the Phe*-Phe* structures ITOBIE and ITOBOK, with high Z'-values, are unique to these structures). Alternating light blue and gold sets of refcodes identify groups of essentially identical structures (redeterminations or pseudopolymorphs differing only in the nature of a smaller, co-crystallized molecule). Additional amino acid abbreviations: Abu = aminobutyric acid, Nva = norvaline, nGly = naphthylglycine. With the exception of the last entry, SBLCYS, L-Xaa-L-Xaa amino acids have absolute chirality SS. In the Comment column chirality is indicated if different from this.

Table S3

Hydrogen bonding patterns in all dipeptide structures

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

Refcode ^a	Seq. ^b	Stereo ^c	Solvent/guest/anion	Packing ^d
<i>N</i> = 0 (15)				
BEQJAJ	IW		dihydrate	T5**
BEQJEN	WV		dihydrate	S5**
BUDXUT	PE		dihydrate	T5**
CAXNUK	RE		dihydrate	no layer, <i>Z'</i> = 2
CELTAO10	YF		hydrate	T5**
CIHNUC	YV		hydrate	T5** = CELTAO10
ETONIK	IF		dihydrate	T5** = BEQJAJ
GASWEC	PK+		acetate	no layer
KIXCAW	YW		hydrate	T5** = CELTAO10
LAMDEK	HS		3.7 hydrate	T4**
MOBYAD	VF		dihydrate	T5** = BEQJAJ
RAVZEU	GH		dihydrate	T5**
VEVGOS	KE		dihydrate	no layer
VIKWUJ	nVF		dihydrate	T5** = BEQJAJ
VUZBAT	RD		dihydrate	no layer
VUZBIB	YL		hydrate	T5** = CELTAO10
ZEFZAL10	YY		dihydrate	no layer
<i>N</i> = 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		(<i>S</i>)-methyl lactate clathrate	anti, <i>Z'</i> = 2
MAVTIM	nGfG		(<i>S</i>)-ethyl pantolactone dimethanol clathr. hydrate	anti = MAVTAE, <i>Z'</i> = 2
MOBYEH	VF		trihydrate	complex, <i>Z</i> = 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
UPUVOR	TfG		1-propanol solvate	anti, mimics GLDLPA
UPUVUX	TfG		butan-2-ol solvate	anti, mimics GLDLPA
UPUWAE	TfG		2-propanol solvate	anti, mimics GLDLPA
UPUWEI	TfG		cyclohexanol solvate	anti, mimics GLDLPA
YAGZOW	II		dihydrate	(S5*)
<i>N</i> = 2 (92)				
ALAMET01	AM	DL/LD		S4

AQAR0Z	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-t , $Z' = 2$
EYIVAJ	SV		T5
FABYEM10	AnV	DL/LD	S4
FEFYUM	LG	0.67 hydrate	unique translation, $Z' = 3$
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-t
GAGFIG	nVV		VA-class
GAGFOM	nVL		VA-class
GAGFUS	nVI		VA-class
GAGPAI	BB		VA-class
GAGPEM	BnV		VA-class
GAGPIQ	BV		VA-class
GAGPOW	BL		VA-class
GAGPUC	BI		VA-class
GAGQAJ	nVB		VA-class
GAGQEN	nVnV		VA-class
GEHTAP	FW	0.75 hydrate	T5-t , $Z' = 4$
GLUGLY	EG		S5
GLYGLY01	GG		T4m (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	<i>SR</i> dimethylformamide clathrate	S4
HUZVUT	fGfG	<i>SR</i> acetamide clathrate	S4
IDUZOW	LL	0.87 hydrate	T5-t , $Z' = 2$
IDUZUC	LF	0.86 hydrate	T5-t , $Z' = 2$
IFABAS	FL	1.26 hydrate	T5-t , $Z' = 2$
IFABEW	FF	2.47 hydrate	T5-t
IKOMOM	LI	trifluoroethanol	S5
ITOBEA	F*F*		S5 , $Z' = 3$
ITOBIE	F*F*	0.5 trifluoroacetate	anti, $Z' = 2$
ITOBOK	F*F*	0.25 trifluoroacetate 0.25 H ₃ O ⁺ 0.125 H ₂ O	anti, $Z' = 2$

JENTOL01	fGfG	(S)-isopropylphenylsulfoxide clathrate	S4
JOQLIM	FF	2.0 methanol	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		S4m
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 (<i>R,R</i>)-bis((<i>o</i> -methylsulfinyl)benzyl)ether clathr.	S4
OKOZAQ	IfG	methyl phenyl sulfoxide	S4
OKOZEU	LfG	methyl 3,4-dimethylphenyl sulfoxide	S4
OKOZIY	IfG	methyl <i>o</i> -tolyl sulfoxide	S5
OLOGEB	MA		VA-class, $Z' = 7$
PAJFIQ	SF		T5
PAJPUM	FI	0.88 hydrate	S5/T5 hybrid, $Z' = 2$
PR SARH	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
TELV OV01	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
This work	VL	acetonitrile	S5

$N = 3$ (25)^h

ALAALA	AA		T5-t
ALAGLY	AG		S5
EMIPAR	AM	0.5 hydrate	S4/T5 hybrid, $Z' = 2$
ETIWIN	LI	0.75 hydrate	no layer, $Z' = 4$
EOVAN	AT		S5m
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			T5m , tubular
JAYFOF	AB			T5/T5/T4 hybrid, $Z' = 3$
LALLSE	AS			S4
MAPKOE	TA			T4-t
MAZYAO	FS	LD		T5m
NIKPOO	BS			T5-t
QQQEVJ01	GF			T5
SEHGES	GP		0.5 hydrate	unique, $Z' = 2$
SERGLY	SG			T4m
WEVWOK	GV			T5 , $Z' = 7$
WILTIV	fGfG	SR		T5m
XEGHOG	GL	L/D		unique
XOSHOC	BA			T5-t
XOSHUI	AB		0.33 hydrate	T5/T5/T4 hybrid, $Z' = 3$

^a Refcode in the CSD (Groom *et al.*, 2016), entries in bold are new since the head-to-tail review (Görlitz, 2010).

^b Peptide sequence given with one-letter amino acid abbreviations. Uncommon acids: nG = (1-naphthyl)glycine; fG = phenylglycine; nV = norvaline; B = 2-amino butyric acid; hP = hydroxyproline; bC = *S*-benzylcysteine; Sar = sarcosine; F* = phenylalanine with-CH₂- replaced by longer chain. An additional '+', as in RS⁺, indicates a peptide net charge of +1, bold typeface indicates a double zwitterion.

^c Stereochemistry is given when deviating from LL, DD, L or D. L/D or DL/LD indicate racemates. An L-amino acid usually has absolute chirality *S*.

^d The basic pattern is given when present. A succeeding 'm', as in **T5m**, indicates a modified pattern. One or two succeeding stars indicates one or two missing C(8) chains, 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.

^h When $Z' > 1$, the high value for *N* may not apply to all molecules.

Table S4

Torsion angles (°) in dipeptide structures for the main structural classes

	CSD refcode and torsion angle as defined in Scheme 1
	Inverted from D-Xaa-D-Xaa structure
	Extreme value within group
	Average value and sample standard deviation
ABCDEF	Representative molecule in group (Fig. 7)

S4

Refcode	ψ_1	ω_1	ϕ_2	ψ_T
COCGIK	132,4	173,0	-150,9	20,0
COCGIK	121,9	172,5	-160,0	-11,6
COCGIK	130,9	173,8	-149,4	19,5
COCGIK	122,3	172,9	-159,3	-8,9
FEHPEO	136,2	178,1	-159,0	-14,1
FEHPAK	129,3	177,9	-162,6	-17,4
COCGOQ	127,9	175,3	-152,6	17,5
COCGOQ	124,9	176,0	-159,0	-5,9
JOQLIM	162,4	170,3	-153,4	14,5
CAQTOD	131,8	174,5	-158,5	-13,2
CAQTOD	141,3	171,5	-140,1	11,6
METMET	133,1	179,9	-149,0	-10,3
OKOZAQ	153,8	178,8	-166,1	-15,8
OKOZEU	144,9	184,8	-159,6	-24,8
DABQOM	131,5	175,1	-146,2	-32,5
JENTOL01	152,4	169,9	-145,5	-19,9
NOTSEU	148,1	168,0	-140,2	-21,7
NOTSIY	145,1	184,9	-159,8	-18,6
LALLSE	124,8	182,0	-156,9	-1,6
BOFZOL	135,5	172,3	-160,3	-18,1
BURLOP	134,3	174,3	-137,7	-32,2
TELVOV01	137,0	175,1	-152,5	-12,6
RUGHIK	165,1	176,7	-168,1	-23,6
XEQQOS	141,3	183,4	-155,0	-29,1
XEQQOS	140,0	178,3	-144,6	-34,1
XEQQEI	143,8	172,8	-146,1	-20,2
XEQQIM	139,1	185,1	-158,6	-18,3
Average	138,2	176,2	-153,7	-11,9
St.dev.	11,2	4,7	8,0	15,9

T4

Refcode	ψ_1	ω_1	ϕ_2	ψ_T
DUHKEW	148,0	161,3	-118,2	-31,9
FOBLUE	134,2	174,7	-160,2	0,4
KIXBID	124,4	187,7	-153,0	-7,4

MAZXUH	136,4	175,3	-162,8	-17,1
MAZYES	145,5	168,1	-164,8	-8,0
MAZYES	132,0	178,3	-162,0	-15,4
RAVZOE	115,3	172,7	-143,4	0,0
Average	133,7	174,0	-152,1	-11,3
St. dev.	11,4	8,2	16,7	11,3

S5

Refcode	ψ_1	ω_1	ϕ_2	ψ_T
ITOB EA	168,6	172,7	-72,9	-33,7
ITOB EA	168,7	166,7	-70,0	-22,7
ITOB EA	168,1	163,6	-71,5	-31,8
BURLIJ	175,8	175,9	-113,6	71,3
CIJGUX	124,5	175,2	-81,7	-30,7
COCGEG	159,9	171,4	-77,6	-19,5
GLUGLY	166,1	175,6	-74,2	-27,1
HIQWAF	168,5	174,9	-127,0	-45,5
HIQWAF	128,8	168,4	-86,1	-29,7
IKOMOM	117,8	173,2	-65,5	-41,1
IKOMOM	162,6	168,6	-99,5	-52,9
JUCSEF01	156,5	179,4	-126,7	-71,6
JUKMEH	163,8	158,8	-110,9	-41,5
OKOZIY	132,4	171,9	-93,0	-22,0
SAMWOT	147,0	170,2	-101,6	-55,4
SAMWUZ	154,8	167,1	-106,1	-43,8
SBLCYS	148,0	174,4	-114,0	87,7
SEGYOS	163,1	172,4	-91,4	-28,9
SEYW AU	129,3	176,5	-90,1	-27,7
SUWLIF	153,0	169,9	-116,5	-64,8
SUWL OL	151,0	162,9	-87,6	-54,8
SUWL OL	151,1	168,1	-85,2	-56,1
SUWL OL	151,5	162,3	-88,4	-54,3
SUWL OL	150,2	168,5	-84,8	-55,2
TARKUT	155,5	172,0	-95,7	-29,4
TEKNAY	156,8	164,9	-109,6	74,2
TIPTOB	141,6	169,2	-101,2	74,9
VIFFEW	161,1	158,6	-106,1	-39,1
XEGNAY	128,9	178,8	-109,6	-80,9
YICGUM	150,4	165,9	-74,5	-27,4
YICGUM	152,2	167,4	-78,8	-26,3
YORPEA	139,1	166,1	-149,5	-14,5
ZILDON	165,2	180,4	-115,3	-41,1
<i>This work</i>	140,9	170,6	-95,9	-11,7
Average	151,6	170,1	-96,2	-25,7
St. dev.	14,5	5,5	19,2	41,3

T5

Refcode	ψ_1	ω_1	ϕ_2	ψ_T
BAPBEZ10	167,5	168,8	-68,0	-22,2
BELCUQ	155,6	167,6	-94,8	-34,4
BIBVOX	167,3	168,5	-73,9	-24,4
BUHGIU	168,2	169,6	-74,6	-18,9
DEZQOO	163,3	168,3	-64,9	-23,7
EYIVAJ	152,6	157,4	-63,7	-24,7
GUKVUD	155,0	158,0	-68,2	-19,1
JUKMOR	164,5	174,8	-78,2	-18,7
PAJFIQ	154,2	157,5	-66,2	-21,2
GLYLEU10	171,6	168,7	-64,9	-30,2
RAVZAQ	160,6	175,0	-77,6	-17,5
SEYWEY	166,3	170,6	-75,8	-22,4
SOJPAI	162,2	174,4	-67,9	-34,9
TIPTUH	156,4	173,7	-75,1	-14,6
QQQEVJ01	166,1	169,2	-66,3	-28,8
Average	162,1	168,1	-72,0	-23,7
St. dev.	6,0	6,0	8,0	6,0

N-term Pro

T5-t

Refcode	ψ_1	ω_1	ϕ_2	ψ_T
ETITUW	136,8	174,6	49,9	47,9
ETITUW	130,9	185,5	50,4	49,3
FULGEY02	156,5	177,0	60,2	31,8
GEHTAP	144,6	181,0	56,0	29,5
GEHTAP	117,9	178,0	51,2	41,8
GEHTAP	106,6	185,8	55,1	33,3
GEHTAP	108,4	188,7	51,2	39,1
IDUZOW	129,6	174,9	47,9	50,6
IDUZOW	127,8	177,9	51,0	51,7
IDUZUC	125,0	179,9	47,6	52,6
IDUZUC	124,4	185,7	49,2	54,1
IFABAS	152,6	183,1	52,5	39,4
IFABAS	155,3	183,0	51,9	44,9
IFABEW	157,8	180,9	55,4	43,7
Average	133,9	181,1	52,1	43,6
St. dev.	17,3	4,4	3,5	8,1

T5**

Refcode	ψ_1	ω_1	ϕ_2	ψ_T
BUDXUT	160,3	173,8	-95,3	-34,9
CELTAO10	164,7	177,4	-70,6	-34,9
CIHNUC	164,6	176,7	-79,3	-45,5
KIXCAW	161,6	177,3	-67,6	-35,2

RAVZEU	164,1	182,2	-103,3	-2,2
VUZBIB	161,8	179,0	-83,5	-35,3
Average	162,8	177,7	-83,3	-31,3
St. dev.	1,8	2,8	13,9	14,9

N-term Tyr

T5**(t)

Refcode	ψ_1	ω_1	ϕ_2	ψ_T
BEQJAJ	147,5	166,0	51,9	49,7
ETONIK	149,9	170,7	49,5	48,3
MOBYAD	151,3	172,3	48,5	48,4
VIKWUJ	166,5	171,5	55,0	37,2
Average	153,8	170,1	51,2	45,9
St. dev.	8,6	2,8	2,9	5,9

S4*

Refcode	ψ_1	ω_1	ϕ_2	ψ_T
ASPGLY	147,101	171,643	152,782	-5,098
KIXBOJ	171,43	169,175	-148,633	6,137
LACBAS	-164,92	175,075	-116,875	-17,001
POTPET02	-174,711	172,753	160,556	-20,944
RAVZUK	159,64	160,749	-71,066	-11,42

References

- Akazome, M., Senda, K. & Ogura, K. (2002). *J. Org. Chem.* **67**, 8885-8889.
- Akazome, M., Ueno, Y., Ooiso, H. & Ogura, K. (2000). *J. Org. Chem.* **65**, 68-76.
- Biswas, A. B., Hughes, E. W., Sharma, B. D. & Wilson, J. N. (1968). *Acta Cryst.* **B24**, 40-50.
- Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Cryst.* **B72**, 171-179.
- Görbitz, C. H. (2001). *Acta Cryst.* **C57**, 575-576.
- Görbitz, C. H. (2004a). *Acta Cryst.* **B60**, 569-577.
- Görbitz, C. H. (2004b). *Acta Cryst.* **C60**, o810-o812.
- Görbitz, C. H. (2010). *Acta Cryst.* **B66**, 84-93.
- Kvick, A., Al-Karaghoul, A. R. & Koetzle, T. F. (1977). *Acta Cryst.* **B33**, 3796-3801.
- Mason, T. O., Chirgadze, D. Y., Levin, A., Adler-Abramovich, L., Gazit, E., Knowles, T. P. J. & Buell, A. K. (2014). *ACS Nano* **8**, 1243-1253.
- Murali, R., Lalitha, V., Subramanian, E. & Parthasarathy, R. (1986). *Int. J. Pept. Protein Res.* **27**, 160-164.