Electronic Supplementary Information:

Crystal Structure of an NPNA-Repeat Motif from the Circumsporozoite Protein of the Malaria Parasite *Plasmodium falciparum*

Arin Ghasparian, Kerstin Moehle, Anthony Linden and John A. Robinson*

Institute of Organic Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

1. NMR studies

^{1}H NMR chemical shift assignments at 600 MHz for the major *trans* rotamer of the peptide Ac-Ala-Asn-Pro-Asn-Ala-NH_2 in H_2O/D_2O (9:1), pH 5, at 278 K were derived by standard methods from DQF-COSY, TOCSY and 250 ms ROESY spectra, and are given below (Table-1S).

Figure-1S. 600 MHz ^{1}H NMR spectrum of Ac-Ala-Asn-Pro-Asn-Ala-NH_2.

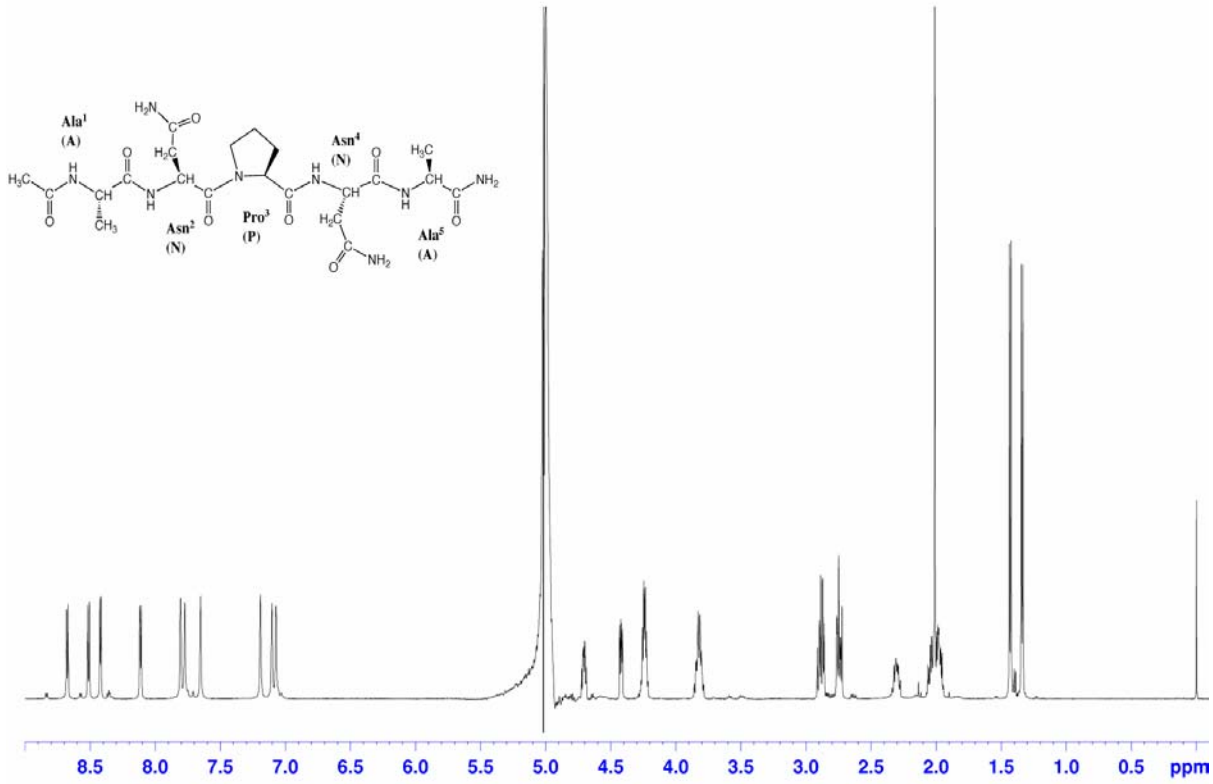


Table-1S. ^{1}H NMR assignments and ^{3}J coupling constants (Hz).

Residue	NH	HC(α)	HC(β)	Others	$^3J_{HNH\alpha}$	$^3J_{H\alpha H\beta}$
Ac-ANPNA-NH$_2$						
Ac	-	-	-	$CH_3(\gamma)$ 2.02		
Ala^1	8.42	4.24	1.34	-	5.5	
Asn^2	8.67	4.97	2.89, 2.74	NH(ε) 7.80; NH(ζ) 7.10	7.3	7.2
Pro^3	-	4.42	2.30, 1.98	$CH_2(\gamma)$ 2.03, 2.03; $CH_2(\delta)$ 3.82, 3.82	-	9.1, 4.5
Asn^4	8.51	4.71	2.88, 2.75	NH(ε) 7.77; NH(ζ) 7.07	7.7	9.1, 5.6
Ala^5	8.11	4.24	1.43	-	5.9	
NH_2	-	-	-	NH(ε) 7.65; NH(ζ) 7.19		

It is not possible here to describe fully the conformational dynamics of this short linear peptide in solution on the basis of the available NMR data. Nevertheless, 2D spectra were analysed in the program XEASY (Bartels, C., Xia, T., Billeter, M, Guentert, P., Wuethrich, K. The programm XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* **1995**, *6*, 1-10). Distance restraints were obtained by integrating cross peaks in ROESY spectra, translating these into upper-distance limits using the CALIBA module of DYANA (Güntert, P.; Mumenthaler, C.; Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **1997**, *273*, 283-298) and correcting for resonance offset effects. The distance restraints are summarized below (Table-2S):

Table-2S. ROE cross peaks and derived distance restraints. The prefix Q indicates the use of a pseudoatom and a pseudoatom correction.

Peak-1		Peak-2		dist.rstrnt (Å)
0 ACE	QAC	1 ALA	HN	4.42
0 ACE	QAC	2 ASN	HN	5.19
1 ALA	HN	1 ALA	HA	3.30
1 ALA	HN	1 ALA	QB	4.82
1 ALA	HN	2 ASN	HN	4.17
1 ALA	HA	2 ASN	HN	2.59
1 ALA	HA	2 ASN	HD22	4.97
1 ALA	QB	2 ASN	HN	6.03
1 ALA	QB	3 PRO	QD	7.42
2 ASN	HN	2 ASN	HA	2.65
2 ASN	HN	2 ASN	HB2	3.30
2 ASN	HN	2 ASN	HB3	3.42
2 ASN	HN	3 PRO	QD	6.39
2 ASN	HN	5 ALA	QB	5.88
2 ASN	HA	2 ASN	HB2	3.89
2 ASN	HA	2 ASN	HB3	4.07
2 ASN	HA	3 PRO	QD	4.63
2 ASN	HB2	2 ASN	HD21	3.79
2 ASN	HB2	2 ASN	HD22	5.50
2 ASN	HB2	3 PRO	QD	6.39
2 ASN	HB3	2 ASN	HD21	4.45
2 ASN	HB3	2 ASN	HD22	5.50
2 ASN	HB3	3 PRO	QD	5.87
2 ASN	HB3	5 ALA	QB	5.47
3 PRO	HA	3 PRO	HB2	3.14
3 PRO	HA	3 PRO	HB3	2.52
3 PRO	HA	3 PRO	QG	5.18
3 PRO	HA	3 PRO	QD	5.68
3 PRO	HA	4 ASN	HN	2.74
3 PRO	HA	5 ALA	HN	4.54
3 PRO	HB2	3 PRO	QD	4.53
3 PRO	HB2	4 ASN	HN	3.39
3 PRO	HB3	3 PRO	QD	5.68
3 PRO	QG	3 PRO	QD	5.52
3 PRO	QD	4 ASN	HN	6.30
4 ASN	HN	4 ASN	HA	3.05
4 ASN	HN	4 ASN	HB2	3.17
4 ASN	HN	4 ASN	HB3	3.67
4 ASN	HN	5 ALA	HN	3.64
4 ASN	HA	4 ASN	HB2	2.96
4 ASN	HA	4 ASN	HB3	2.71
4 ASN	HA	4 ASN	HD21	5.50
4 ASN	HA	4 ASN	HD22	5.50
4 ASN	HA	5 ALA	HN	2.83

4	ASN	HA	6	NH2	HEN	5.50
4	ASN	HB2	4	ASN	HD21	3.92
4	ASN	HB2	4	ASN	HD22	5.50
4	ASN	HB2	5	ALA	HN	4.23
4	ASN	HB3	4	ASN	HD21	4.85
4	ASN	HB3	4	ASN	HD22	5.50
4	ASN	HB3	5	ALA	HN	4.11
4	ASN	HB3	5	ALA	HA	4.35
5	ALA	HN	5	ALA	HA	3.21
5	ALA	HN	5	ALA	QB	4.67
5	ALA	HN	6	NH2	HEN	5.00
5	ALA	HA	6	NH2	HEN	3.58
5	ALA	HA	6	NH2	HZN	4.82
5	ALA	QB	6	NH2	HEN	5.69
5	ALA	QB	6	NH2	HZN	5.88

2. Crystallography

The structure of $C_{21}H_{34}N_8O_8 \cdot 3H_2O$ (AGI62) has been solved and refined successfully with no unusual features. The space group permits the compound in the crystal to be enantiomerically pure, but the absolute configuration of the molecule has not been determined. The enantiomer used in the refinement was based on the expected *S*-configuration of all peptide units in the molecule. The asymmetric unit contains one peptide and three water molecules. The peptide molecule displays two intramolecular hydrogen bonds, which serve to maintain a turn conformation. The remaining N-H groups and water molecules are involved in intermolecular hydrogen bonds which link the peptide and water molecules into an infinite three-dimensional framework. See Table 3S for full details.

Figure-2S. *ORTEP*[1] representation of the molecule (50% probability ellipsoids; H-atoms given arbitrary displacement parameters for clarity)

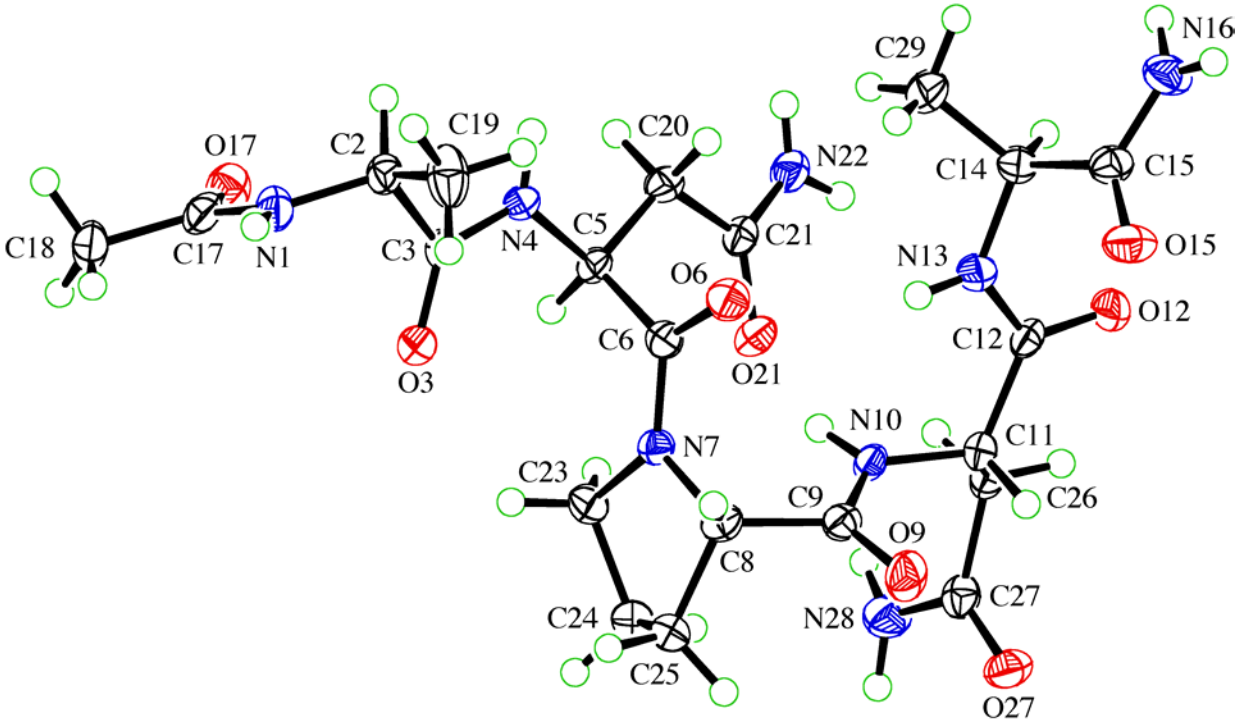
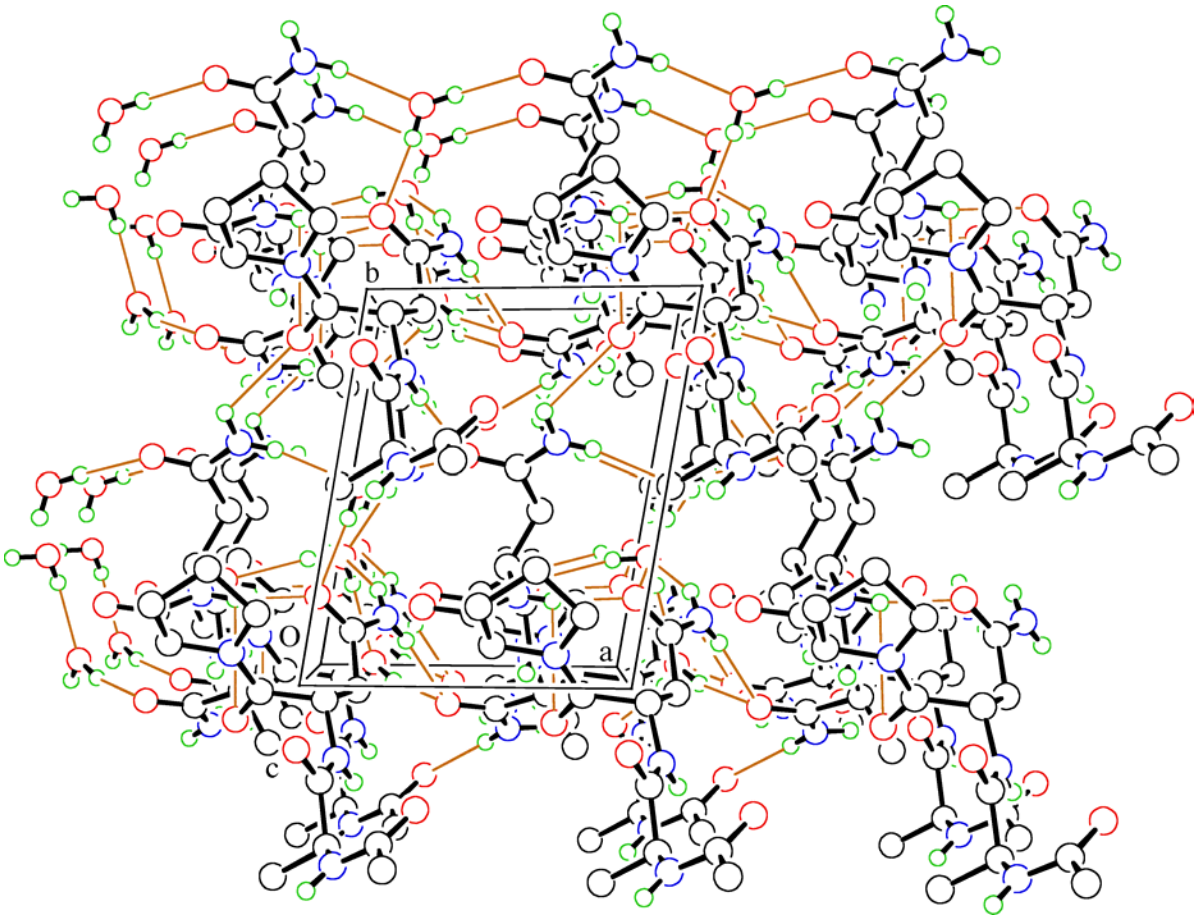


Figure-3S. Molecular packing projected down the *c*-axis showing the hydrogen bonding scheme (equivalent isotropic spheres for atoms; uninvolved H-atoms omitted for clarity).



EXPERIMENTAL

Crystal-Structure Determination. – A crystal of $C_{21}H_{34}N_8O_8 \cdot 3H_2O$, obtained from D_2O, was mounted on a glass fibre and used for a low-temperature X-ray structure determination. All measurements were made on a *Nonius KappaCCD* area-detector diffractometer[2] using graphite-monochromated Mo *K*α radiation (λ = 0.71073 Å) and an *Oxford Cryosystems Cryostream 700* cooler. The unit cell constants and an orientation matrix for data collection were obtained from a least-squares refinement of the setting angles of 2373 reflections in the range $4° < 2\theta < 50°$. The mosaicity was 0.527(3)°. A total of 208 frames were collected using ω scans with κ offsets, 10 seconds exposure time and a rotation angle of 2.0° per frame, and a crystal-detector distance of 30.0 mm.

Data reduction was performed with *HKL Denzo* and *Scalepack*[3]. The intensities were corrected for Lorentz and polarization effects, but not for absorption. The space group was determined from packing considerations, a statistical analysis of intensity distribution, and the

successful solution and refinement of the structure. Equivalent reflections were merged. Data collection and refinement parameters are given in the cif file. A view of the molecule is shown in Figure-2S. The crystallographic data are available through the Cambridge Crystallographic Data Center (CCDC 280279).

The structure was solved by direct methods using *SIR92*[4], which revealed the positions of all non-hydrogen atoms. The asymmetric unit contains one peptide and three water molecules. The non-hydrogen atoms were refined anisotropically. The H-atoms of the water molecules and the primary amide N-atoms were placed in the positions indicated by a difference electron density map and their positions were allowed to refine together with individual isotropic displacement parameters. All remaining H-atoms were placed in geometrically calculated positions and refined using a riding model where each H-atom was assigned a fixed isotropic displacement parameter with a value equal to $1.2U_{eq}$ of its parent C-atom ($1.5U_{eq}$ for the methyl groups). Refinement of the structure was carried out on F^2 using full-matrix least-squares procedures, which minimised the function $\Sigma w(F_o^2 - F_c^2)^2$. The weighting scheme was based on counting statistics and included a factor to downweight the intense reflections. Plots of $\Sigma w(F_o^2 - F_c^2)^2$ versus $F_c / F_c(\max)$ and resolution showed no unusual trends. A correction for secondary extinction was not applied. One reflection, whose intensity was considered to be an extreme outlier, was omitted from the final refinement.

Neutral atom scattering factors for non-hydrogen atoms were taken from Maslen, Fox and O'Keefe[5a], and the scattering factors for H-atoms were taken from Stewart, Davidson and Simpson[6]. Anomalous dispersion effects were included in F_c[7]; the values for f' and f'' were those of Creagh and McAuley[5b]. The values of the mass attenuation coefficients are those of Creagh and Hubbel[5c]. All calculations were performed using the *SHELXL97*[8] program.

Table-3S. Hydrogen bonding geometry (Å, °).

D	H	A	D-H	H···A	D···A	D-H···A
N(1)	-H(1)	···O(30^{i})	0.88	2.10	2.954(4)	164
N(4)	-H(4)	···O(27^{ii})	0.88	1.99	2.863(4)	172
N(10)	-H(10)	···O(21)	0.88	1.94	2.774(3)	157
N(13)	-H(13)	···O(6)	0.88	2.43	3.257(4)	157
N(16)	-H(161)	···O(3^{iii})	0.94(5)	1.89(5)	2.837(4)	178(4)
N(16)	-H(162)	···O(17^{iv})	0.83(4)	2.05(4)	2.873(4)	173(6)
N(22)	-H(221)	···O(15^{v})	0.83(5)	1.97(5)	2.790(4)	171(4)
N(22)	-H(222)	···O(30^{v})	0.80(6)	2.44(6)	3.128(5)	144(5)
N(28)	-H(281)	···O(6^{vi})	0.91(5)	2.50(5)	2.952(4)	111(3)
N(28)	-H(282)	···O(31^{v})	0.86(5)	2.12(5)	2.971(5)	168(4)
O(30)	-H(301)	···O(12^{vii})	0.88(6)	2.10(6)	2.969(4)	168(5)

O(30) -H(302)···O(32)	0.8(1)	2.0(1)	2.779(5)	154(8)
O(31) -H(311)···O(27)	0.82(7)	2.09(7)	2.888(4)	163(6)
O(31) -H(312)···O(21^{vii})	0.81(8)	1.99(8)	2.798(4)	174(7)
O(32) -H(321)···O(15)	0.89(7)	2.07(7)	2.945(4)	165(5)
O(32) -H(322)···O(3^{iv})	0.90(7)	1.99(7)	2.855(4)	161(5)

Atoms with Roman numerals refer to the following symmetry related positions:

i $1+x, -1+y, 1+z$ ii $1+x, -1+y, z$ iii $x, y, -1+z$

iv $-1+x, y, -1+z$ v $1+x, y, z$ vi $x, 1+y, z$

vii $-1+x, y, z$

N(10)-H and N(13)-H form intramolecular hydrogen bonds with the amide O-atoms that are seven atoms back along the peptide backbone. Each of these interactions has the graph set motif[9] of S(10) and serves to maintain a helical conformation at one end of the peptide. The remaining N-H groups are involved in intermolecular hydrogen bonds with amide O-atoms of neighbouring molecules or with O-atoms of the water molecules. The water molecules form intermolecular hydrogen bonds with other water molecules and with amide O-atoms of neighbouring peptide molecules. The intermolecular interactions link the peptide and water molecules into an infinite three-dimensional framework.

REFERENCES

(1) C.K. Johnson, *ORTEPII*, Report ORNL-5138, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1976.

(2) R. Hooft, *KappaCCD Collect Software*, Nonius BV, Delft, The Netherlands, 1999.

(3) Z. Otwinowski, W. Minor, in *Methods in Enzymology*, Vol. 276, *Macromolecular Crystallography*, Part A, Eds. C.W. Carter Jr., R.M. Sweet, Academic Press, New York, 1997, pp. 307-326.

(4) A. Altomare, G. Cascarano, C. Giacovazzo, A. Guagliardi, M.C. Burla, G. Polidori, M. Camalli, *SIR92*, *J. Appl. Crystallogr.* **1994**, *27*, 435.

(5) a) E.N. Maslen, A.G. Fox, M.A. O'Keefe, in 'International Tables for Crystallography', Ed. A.J.C. Wilson, Kluwer Academic Publishers, Dordrecht, 1992, Vol. C, Table 6.1.1.1,

pp. 477-486; b) D.C. Creagh, W.J. McAuley, *ibid*. Table 4.2.6.8, pp. 219-222; c) D.C. Creagh, J.H. Hubbell, *ibid*. Table 4.2.4.3, pp. 200-206.

(6) R.F. Stewart, E.R. Davidson, W.T. Simpson, *J. Chem. Phys.* **1965**, *42*, 3175-3187.

(7) J.A. Ibers, W.C. Hamilton, *Acta Crystallogr.* **1964**, *17*, 781-782.

(8) G.M. Sheldrick, *SHELXL97, Program for the Refinement of Crystal Structures*, University of Göttingen, Germany, 1997.

(9) J. Bernstein, R.E. Davis, L. Shimoni, N.-L. Chang, *Angew. Chem.* **1995**, *107*, 1689-1708; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 1555-1573.

Coordinates in PDB format:

```
ATOM      1 O3           0       9.530  -1.192  10.523  1.00  0.00           O
ATOM      2 O6           0       7.288  -0.792   7.358  1.00  0.00           O
ATOM      3 O9           0       4.271   1.903   7.753  1.00  0.00           O
ATOM      4 O12          0       5.959   2.277   3.194  1.00  0.00           O
ATOM      5 O15          0       3.988  -0.689   3.452  1.00  0.00           O
ATOM      6 O17          0      12.356  -2.467  10.932  1.00  0.00           O
ATOM      7 O21          0       9.067   2.063   6.451  1.00  0.00           O
ATOM      8 O27          0       5.092   5.364   6.597  1.00  0.00           O
ATOM     12 N1           0      10.520  -3.718  11.211  1.00  0.00           N
ATOM     13 N4           0       9.938  -1.608   8.347  1.00  0.00           N
ATOM     14 N7           0       7.592   0.872   8.850  1.00  0.00           N
ATOM     15 N10          0       6.314   2.135   6.790  1.00  0.00           N
ATOM     16 N13          0       6.243   0.506   4.559  1.00  0.00           N
ATOM     17 N16          0       5.141  -1.422   1.668  1.00  0.00           N
ATOM     18 N22          0      10.322   1.466   4.690  1.00  0.00           N
ATOM     19 N28          0       7.304   5.690   6.925  1.00  0.00           N
ATOM     20 C2           0      10.063  -3.469   9.865  1.00  0.00           C
ATOM     21 C3           0       9.848  -1.974   9.618  1.00  0.00           C
ATOM     22 C5           0       9.574  -0.263   7.898  1.00  0.00           C
ATOM     23 C6           0       8.058  -0.078   8.027  1.00  0.00           C
ATOM     24 C8           0       6.148   1.106   9.006  1.00  0.00           C
ATOM     25 C9           0       5.495   1.755   7.780  1.00  0.00           C
ATOM     26 C11          0       5.871   2.738   5.551  1.00  0.00           C
ATOM     27 C12          0       6.036   1.820   4.337  1.00  0.00           C
ATOM     28 C14          0       6.407  -0.417   3.454  1.00  0.00           C
ATOM     29 C15          0       5.062  -0.847   2.847  1.00  0.00           C
ATOM     30 C17          0      11.668  -3.191  11.658  1.00  0.00           C
ATOM     31 C18          0      12.061  -3.563  13.055  1.00  0.00           C
ATOM     32 C19          0       8.742  -4.188   9.636  1.00  0.00           C
ATOM     33 C20          0      10.020  -0.139   6.446  1.00  0.00           C
ATOM     34 C21          0       9.775   1.224   5.852  1.00  0.00           C
ATOM     35 C23          0       8.396   1.828   9.642  1.00  0.00           C
ATOM     36 C24          0       7.371   2.863  10.090  1.00  0.00           C
ATOM     37 C25          0       6.080   2.048  10.220  1.00  0.00           C
ATOM     38 C26          0       6.609   4.057   5.293  1.00  0.00           C
ATOM     39 C27          0       6.273   5.091   6.337  1.00  0.00           C
ATOM     40 C29          0       7.190  -1.649   3.891  1.00  0.00           C
ATOM     41 H1           0      10.037  -4.224  11.745  1.00  0.00           H
ATOM     42 H2           0      10.742  -3.814   9.217  1.00  0.00           H
ATOM     43 H4           0      10.226  -2.192   7.755  1.00  0.00           H
ATOM     44 H5           0      10.047   0.419   8.456  1.00  0.00           H
ATOM     45 H8           0       5.688   0.245   9.220  1.00  0.00           H
ATOM     46 H10          0       7.178   2.011   6.901  1.00  0.00           H
ATOM     47 H11          0       4.897   2.947   5.645  1.00  0.00           H
ATOM     48 H13          0       6.280   0.206   5.385  1.00  0.00           H
```

```
ATOM     49 H14          0       6.934   0.046   2.740  1.00  0.00           H
ATOM     50 H161         0       5.940  -1.428   1.166  1.00  0.00           H
ATOM     51 H162         0       4.498  -1.809   1.300  1.00  0.00           H
ATOM     52 H181         0      12.458  -2.787  13.501  1.00  0.00           H
ATOM     53 H182         0      12.713  -4.294  13.028  1.00  0.00           H
ATOM     54 H183         0      11.267  -3.853  13.552  1.00  0.00           H
ATOM     55 H191         0       8.073  -3.851  10.267  1.00  0.00           H
ATOM     56 H192         0       8.866  -5.150   9.774  1.00  0.00           H
ATOM     57 H193         0       8.436  -4.026   8.718  1.00  0.00           H
ATOM     58 H201         0      10.986  -0.343   6.390  1.00  0.00           H
ATOM     59 H202         0       9.537  -0.814   5.905  1.00  0.00           H
ATOM     60 H221         0      10.732   0.886   4.261  1.00  0.00           H
ATOM     61 H222         0      10.055   2.110   4.292  1.00  0.00           H
ATOM     62 H231         0       9.103   2.243   9.087  1.00  0.00           H
ATOM     63 H232         0       8.817   1.382  10.420  1.00  0.00           H
ATOM     64 H241         0       7.272   3.582   9.418  1.00  0.00           H
ATOM     65 H242         0       7.627   3.266  10.958  1.00  0.00           H
ATOM     66 H251         0       5.282   2.632  10.177  1.00  0.00           H
ATOM     67 H252         0       6.063   1.539  11.069  1.00  0.00           H
ATOM     68 H261         0       7.585   3.893   5.297  1.00  0.00           H
ATOM     69 H262         0       6.361   4.402   4.399  1.00  0.00           H
ATOM     70 H281         0       7.109   6.390   7.479  1.00  0.00           H
ATOM     71 H282         0       8.090   5.512   6.613  1.00  0.00           H
ATOM     72 H291         0       6.707  -2.103   4.613  1.00  0.00           H
ATOM     73 H292         0       8.076  -1.377   4.212  1.00  0.00           H
ATOM     74 H293         0       7.291  -2.259   3.131  1.00  0.00           H
ATOM      9 O30          0       1.151   2.826   2.350  1.00  0.00           O
ATOM     75 H301         0       0.316   2.785   2.641  1.00  0.00           H
ATOM     76 H302         0       1.281   2.134   1.919  1.00  0.00           H
ATOM     10 O31          0       2.347   4.683   6.014  1.00  0.00           O
ATOM     77 H311         0       3.136   4.918   6.014  1.00  0.00           H
ATOM     78 H312         0       2.150   3.903   6.149  1.00  0.00           H
ATOM     11 O32          0       1.757   0.183   1.739  1.00  0.00           O
ATOM     79 H321         0       2.372  -0.233   2.239  1.00  0.00           H
ATOM     80 H322         0       1.369  -0.384   1.155  1.00  0.00           H
```