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*

Received (in Cambridge, UK) 29th July 2005, Accepted 18th October 2005
First published as an Advance Article on the web 21st November 2005
DOI: 10.1039/b510812h

A crystal structure is reported of the peptide Ac-Ala-Asn-Pro-Asn-Ala-NH_2, representing the immunodominant region of the major surface protein on the malaria parasite; the NPNA motif adopts a type-I β-turn, which is stabilized by hydrogen bonding between the CO of Asn^2 and the NH of Ala^5 as well as between the O(δ) of Asn^2 and the NH of Asn^4.

Malaria, the most important parasitic infection in humans, is caused by protozoan parasites of the genus *Plasmodium*, of which *P. falciparum* causes the majority of infections in Africa and is responsible for the most severe form of the disease.[1] The major circumsporozoite (CS) surface protein has been the focus of numerous efforts over many years to develop a pre-erythrocyte vaccine that would prevent infection of hepatocytes by sporozoites.[2] The CS protein forms a dense coat covering the entire surface of sporozoites, it mediates sporozoite adhesion to target cells, and is required for sporozoite development in the mosquito.[3] The CS proteins from all species of *Plasmodium* show a similar overall size (*ca.* 400 amino acids) and domain organization, including a signal peptide, a central region composed mostly of amino acid repeat motifs, and a C-terminal GPI membrane anchor. The repeat motifs in the *P. falciparum* CS protein comprise 37 tandem repeats of the tetrapeptide Asn-Ala-Asn-Pro (NANP) interspersed with four copies of Asn-Val-Asp-Pro (NVDP). The function of the CS repeats is unknown, but it has been speculated that they may fulfil a structural role, with repeats from neighbouring CS molecules interlocked to form a sheath surrounding the parasite.[3]

Antibody responses to the CS protein are primarily directed against immunodominant epitopes in the central NANP (or less commonly quoted as NPNA) repeat region. $(NANP)_3$ peptides formed the basis of the first synthetic and recombinant subunit malaria vaccine candidates tested in human clinical trials.[4] Moreover, the current leading malaria vaccine candidate, for which clinical development is most advanced, called **RTS,S/AS02A** (GlaxoSmithKline Biologicals) is based on two recombinant proteins, one (**RTS**) containing amino acids 207–395 of the CS protein, which includes 19 copies of the NANP repeat, fused to the amino terminus of the hepatitis B surface antigen (HBsAg), and the second (**S**) being the 226 amino acid HBsAg, formulated together with **AS02A** adjuvant.[5]

Institute of Organic Chemistry, University of Zurich, Winterthurerstrasse 190, 8057, Zurich, Switzerland. E-mail: robinson@oci.unizh.ch

† Electronic supplementary information (ESI) available: ^{1}H NMR assignments and distance restraints, and crystallographic data. See DOI: 10.1039/b510812h

Presently no crystal structure of the CS protein of any *Plasmodium* species has been reported. Earlier NMR and CD studies of synthetic peptides based on the NANP and NPNA cadences, and composed of one to three tetrapeptide motifs, provided evidence for the presence of turn-like structures based on NPNA, stabilized by hydrogen bonding and in rapid dynamic equilibrium with extended chain forms in aqueous solution.[6] NMR studies on $(NANP)_x$ in water/methanol mixtures also suggested that fast conformational averaging occurs together with mixed β-turns and half turns and/or γ-turns,[7] and similar conclusions were reached in studies performed in DMSO.[8] The $(NANP)_3$ epitope has also been studied by multidimensional heteronuclear NMR in lipid micelles[9] in the context of an N-terminal insert into the pVIII protein of bacteriophage fd. In this case, two extended non-hydrogen-bonded NPNA loops were found, with N3 and N5 in the β-region and A6, N7, N9 and A10 in the α-region of ϕ/ψ space. Several attempts have also been made to predict the structure of the repeated tetrapeptide motif using computational methods. Brooks *et al.* postulated an extended right-handed 12_{38} helix for a repeating NPNA oligomer,[10] whereas Gibson and Scheraga found left-handed and right-handed helical conformations.[11]

In earlier work, several conformationally constrained NPNA-mimetics were synthesized and their conformations in water were examined by NMR and CD methods. Thus, substituting proline for (*S*)-α-methylproline (P^{Me}) stabilized a hydrogen bonded type-I β-turn within the NPNA motif in linear $(NP^{Me}NA)_n$ peptides.[12] More recently, conformational studies of cyclic peptidomimetics led to the suggestion that repeating NPNA oligomers might prefer a conformation in which the first Asn in each NPNA module is in the β-region of ϕ/ψ space and the next three (Pro-Asn-Ala) are each in the helical region, thus forming a so-called helical-turn.[13,14] The proposed model of an NPNA oligomer is then not a continuous helix, but the NPNA modules can form periodic structures stabilized by backbone–backbone and backbone–side-chain hydrogen bonds.

During new NMR studies, the peptide Ac-Ala-Asn-Pro-Asn-Ala-NH_2 (Fig. 1) crystallized from H_2O/D_2O (9 : 1) solution (pH 5) over an extended period in the NMR tube. We report here the crystal structure and some relevant NMR properties of this peptide‡. This is the first reported crystal structure of any fragment of the CS protein of *P. falciparum*.

The crystal structure is characterized by a type I β-turn with residues Pro^3 at position $i + 1$ and Asn^4 at position $i + 2$ and backbone torsion angles $\phi_{i+1} = -71.3(4)^\circ$, $\psi_{i+1} = -5.9(4)^\circ$, $\phi_{i+2} = -109.9(3)^\circ$ and $\psi_{i+2} = +15.7(4)^\circ$ (Fig. 2). A further common feature of β turns, the formation of a hydrogen bond between

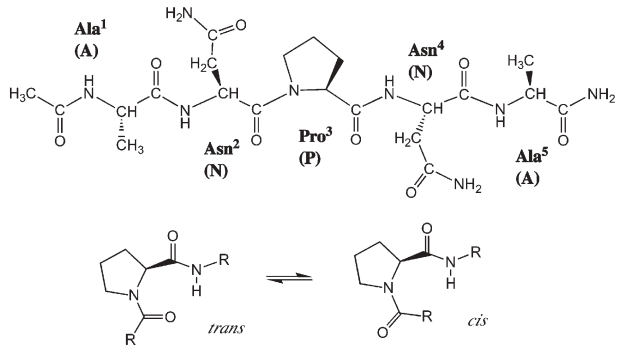

Fig. 1 The peptide Ac-ANPNA-NH_2 and *cis–trans* rotamers at Asn-Pro.

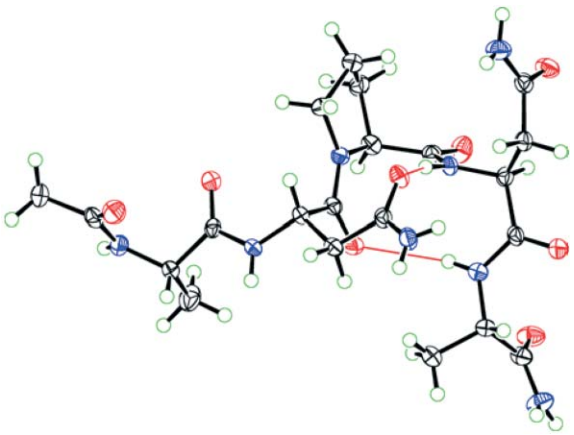

Fig. 2 X-ray crystal structure of Ac-ANPNA-NH_2.

C=O of residue *i* and NH of residue *i* + 3, is observed between Asn^2 and Ala^5 ($d_{H...O}$ = 2.43 Å, $\angle_{NH..O}$ = 157°). The Asn residue preceding Pro adopts an extended conformation with backbone torsion angles ϕ = −68.7(4)° and ψ = +118.1(3)°. An additional intramolecular hydrogen bond is formed between the side chain O^{δ} of Asn^2 and the backbone NH group of Asn^4 ($d_{H...O}$ = 1.94 Å, $\angle_{NH..O}$ = 157°) which could exert a further stabilizing effect on the type I β-turn geometry. The Ala residues representing the start and end residues of ANPNA adopt extended conformations with torsion angles ϕ = −60.2(4)° and ψ = +155.2(3)° for Ala^1 and ϕ = −81.2(4)° and ψ = +164.8(3)° for Ala^5. These angles could be influenced by crystal packing effects and/or intermolecular hydrogen bond formation. The pyrrolidine ring of proline adopts the Cγ-endo (or DOWN) puckered conformation (χ_1 (N–Cα–Cβ–Cγ) = +32.6(3)°; χ_2 (Cα–Cβ–Cγ–Cδ) = −39.8(3)°).

In aqueous solution by ^{1}H NMR spectroscopy at 278 K, the peptide shows two conformational isomers on the NMR time-scale, which can be assigned from ROESY spectra to *trans* : *cis* rotamers (Fig. 1) in 93 : 7 ratio at the Asn-Pro peptide bond (see ESI). ROESY cross peaks and ^3J coupling constants observed for the peptide generally support the earlier conclusions of Dyson and co-workers with related peptides, in showing that β-turn structures are populated by this peptide in aqueous solution. However, the observation of a strong Pro^3(Hα)-Asn^4(HN) as well as a weak Pro^3(Hδ)-Asn^4(HN) ROE suggests that both type-I and type-II β-turns are present in solution, most likely in rapid dynamic equilibrium with extended forms (see ESI).

The conformation of the NPNA motif in the crystal structure need not necessarily reflect the preferred conformation(s) that exist in the native CS protein on the parasite. However, assuming that the crystal structure is relevant, it is interesting, using this information, to consider how a longer NPNA-oligomer might fold. In this respect, the conformation in the crystal structure is very similar (Fig. 3B) to that reported earlier for the $NP^{Me}NA$ motif in $(NP^{Me}NA)_n$ peptides studied in water by NMR,[12] but is quite unlike the conformations of NPNA motifs in NPNA oligomers deduced earlier by computer modeling.[10,11] Also, the helical turn conformation for NPNA discussed above, with ϕ/ψ for Asn^1 in the β-region and Pro^2-Ala^4 in the α-region, which might

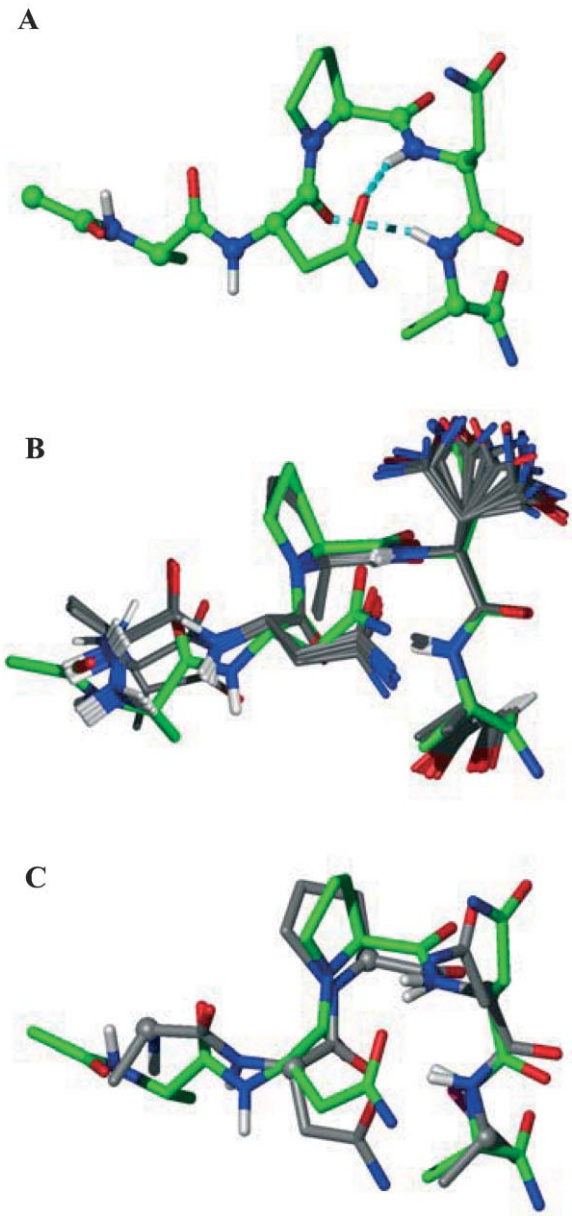

Fig. 3 **A**, Stick representation of the crystal structure in green/blue/red (compare Fig. 2) superimposed upon; **B**, average NMR structures for $(NP^{Me}NA)_n$ reported earlier;[12] and **C**, a computer model of ANPNA in a helical turn conformation[14] (ϕ/ψ for N^2 = −123°/+70°; P^3–N^4 = −60°/−30°; A^5 = −90°/−30°).

represent a conformational motif repeated in the native CS protein,[14] is very closely related to the crystal structure of the peptide, as shown in Fig. 3C.

Finally, the results show conclusively that a conformational repeat is preferred, based on a turn, in the NPNA cadence rather than the more commonly quoted NANP sequence. Furthermore, the crystal structure should be helpful in studying the mechanisms of antigen recognition by inhibitory anti-NPNA-repeat antibodies, and perhaps also in the design of more effective, conformationally defined malaria vaccine candidates.

This work was supported by the Swiss Commission for Technology and Innovation (CTI) and the Swiss National Science Foundation.

Notes and references

‡ Crystal structure data: Crystals from H_2O/D_2O, $C_{21}H_{34}N_8O_8{\cdot}3H_2O$, M_r = 580.59, triclinic, space group $P1$, a = 7.6005(3), b = 9.0722(4), c = 10.5034(5) Å, α = 86.798(3), β = 79.255(3), γ = 79.767(3)°, V = 700.03(5) Å^3, Z = 1, d_{calc} = 1.377 g cm^{-3}, μ(Mo-Kα) = 0.111 mm^{-1}, T = 160 K, 9712 reflections measured, 2415 unique (R_{int} = 0.068), 2257 with $I > 2\sigma(I)$, refinement on F^2 with SHELXL-97, 412 parameters, $R(F)$ [$I > 2\sigma(I)$ reflections] = 0.044, $wR(F^2)$ (all data) = 0.106. Absolute configuration not determined, but assigned from the known *S*-absolute configuration of the amino acids used in the synthesis. CCDC 280279. For crystallographic data in CIF or other electronic format see DOI: 10.1039/b510812h

1 B. M. Greenwood, K. Bojang, C. J. M. Whitty and G. A. T. Targett, *Lancet*, 2005, **365**, 1487.

2 V. S. Moorthy, M. F. Good and A. V. S. Hill, *Lancet*, 2004, **363**, 150; W. R. Ballou, M. Arevalo-Herrera, D. Carucci, T. L. Richie, G. Corradin, C. Diggs, P. Druilhe, B. K. Giersing, A. Saul, D. G. Heppner, K. E. Kester, D. E. Lanar, J. Lyon, A. V. S. Hill, W. Pan and J. D. Cohen, *Am. J. Trop. Med. Hyg.*, 2004, **71**, Suppl. 2, 239.

3 S. H. I. Kappe, C. A. Buscaglia and V. Nussenzweig, *Annu. Rev. Cell Dev. Biol.*, 2004, **20**, 29.

4 D. A. Herrington, D. F. Clyde, G. Losonsky, M. Cortesia, J. R. Murphy, J. Davis, S. Baqar, A. M. Felix, E. P. Heimer, D. Gillessen, E. Nardin, R. S. Nussenzweig, V. Nussenzweig, M. R. Hollingdale and M. M. Levine, *Nature*, 1987, **328**, 257; W. R. Ballou, J. A. Sherwood, F. A. Neva, D. M. Gordon, R. A. Wirtz, G. F. Wasserman, C. L. Diggs, S. L. Hoffman, M. R. Hollingdale, W. T. Hockmeyer, I. Schneider, J. F. Young, P. Reeve and J. D. Chulay, *Lancet*, 1987, **1**, 1277.

5 D. G. Heppner, K. E. Kester, C. F. Ockenhouse, N. Tornieporth, O. Ofori, J. A. Lyon, V. A. Stewart, P. Dubois, D. E. Lanar, U. Krzych, P. Moris, E. Angov, J. F. Cummings, A. Leach, B. T. Hall, S. Dutta, R. Schwenk, C. Hillier, A. Barbosa, L. A. Ware, L. Nair, C. A. Darko, M. R. Withers, B. Ogutu, M. E. Polhemus, M. Fukuda, S. Pichyangkul, M. Gettyacamin, C. Diggs, L. Soisson, J. Milman, M.-C. Dubois, N. Garcon, K. Tucker, J. Wittes, C. V. Plowe, M. A. Thera, O. K. Duombo, M. G. Pau, J. Goudsmit, W. R. Ballou and J. Cohen, *Vaccine*, 2005, **23**, 2243.

6 H. J. Dyson, A. C. Satterthwait, R. A. Lerner and P. E. Wright, *Biochemistry*, 1990, **29**, 7828.

7 G. Esposito, A. Pessi and A. S. Verdini, *Biopolymers*, 1989, **28**, 225.

8 K. Umemoto, J. Kikuchi, M. Narita, K. Fujita and T. Asakura, *Polymer J.*, 1995, **27**, 347.

9 M. Monette, S. J. Opella, J. Greenwood, A. E. Willis and R. N. Perham, *Protein Sci.*, 2001, **10**, 1150.

10 B. R. Brooks, R. W. Pastor and F. W. Carson, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 4470.

11 K. D. Gibson and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 5649.

12 C. Bisang, C. Weber, J. Inglis, C. A. Schiffer, W. F. van Gunsteren, I. Jelesarov, H. R. Bosshard and J. A. Robinson, *J. Am. Chem. Soc.*, 1995, **117**, 7904; A. P. Nanzer, A. E. Torda, C. Bisang, C. Weber, J. A. Robinson and W. F. van Gunsteren, *J. Mol. Biol.*, 1997, **267**, 1012.

13 R. Moreno, L. Jiang, K. Moehle, R. Zurbriggen, R. Glück, J. A. Robinson and G. Pluschke, *ChemBioChem*, 2001, **2**, 838.

14 B. Pfeifer, E. Perduzzi, K. Moehle, R. Zurbriggen, R. Glück, G. Pluschke and J. A. Robinson, *Angew. Chem., Int. Ed.*, 2003, **42**, 2368.

