DOI: 10.1002/cmdc.200700236

Exploring the Flap Pocket of the Antimalarial Target Plasmepsin II: The "55% Rule" Applied to Enzymes

Martina Zürcher,[a] Thomas Gottschalk,[a] Solange Meyer,[b] Daniel Bur,[b] and François Diederich*[a]

With 300–660 million infections annually, malaria is a major health issue and threatens approximately 40% of the world's population.[1,2] In the search for new antimalarial targets, attention has turned to the degradation of human hemoglobin by the parasite *Plasmodium falciparum*. Several enzymes, including the three aspartic proteases plasmepsins (PMs) I, II, IV, and a histo-aspartic protease (HAP), are involved in the process, which is both vital and specific for the pathogen. As these proteases show overlapping substrate specificity, all of them have to be targeted by a potential antimalarial drug.[3,4] Unfortunately, to date no crystallographic data are available for PM I or HAP in the protein data bank (PDB). The only two available X-ray crystal structures of PM IV in complex with pepstatin A[5] and an allophenylnorstatin-based inhibitor (PM IV of *Plasmodium malariae*)[6] are showing a pepsinlike fold, which is also observed for the seventeen published X-ray crystal structures of PM II. However, three of these PM II structures (PDB codes 2BJU, 2IGX, 2IGY)[7,8] feature a new cavity (flap pocket). This pocket is opened and shaped by the *n*-pentyl chain of the reported inhibitors.[7] The existence of this pocket has been observed before in renin.[9] It is mainly lined by hydrophobic amino acid residues, and only its occupancy resulted in potent and selective nonpeptidomimetic inhibitors for PMs and HAP.[8,10] Related human aspartic proteases such as cathepsin D and E must not be affected, and indeed the activity of both was only insignificantly impaired.

The exploration of the molecular recognition properties of the plasmepsin proteases, especially with respect to the flap pocket, is a major objective of our research, and in this work we describe the optimal filling of this hydrophobic cavity.

Recently, a new class of nonpeptidomimetic inhibitors targeting the flap pocket was designed and synthesized in our laboratory (Figure 1).[10] Inhibitor (±)-**1** was the most active compound against PM II (IC_{50} = 130 (PM II) and IC_{50} = 50 nM (PM IV); IC_{50} = concentration of inhibitor at which 50% of the maximum initial rate is observed), whereas compound (±)-**2** was the most potent against PM IV (IC_{50} = 210 (PM II) and IC_{50} = 30 nM (PM IV)). Enantiomer (−)-**2** binds much better to PM II (IC_{50} = 45 nM) and PM IV (IC_{50} = 10 nM) than (+)-**2** (IC_{50} = 3260 and 33900 nM, respectively). The absolute configuration of the active enantiomer was tentatively assigned based on molecular modeling.[10]

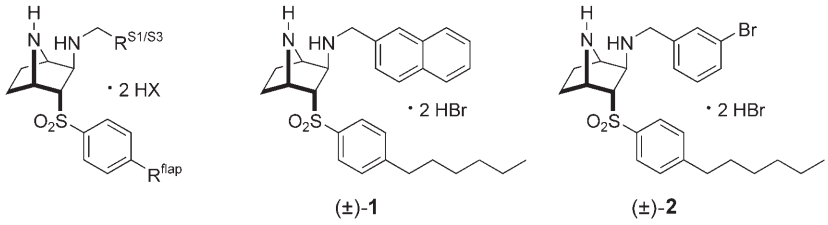

Figure 1. General scaffold of our plasmepsin inhibitors, lead compound (±)-**1**, and inhibitor (±)-**2**. The (protonated) azanorbornane addresses the catalytic Asp dyad, the alkyl chain binds into the flap pocket, and the aromatic residue occupies the S1/S3 site.

We chose compound (±)-**1** as a lead structure. With the residue $R^{S1/S3}$ targeting the S1/S3 subsite left unchanged (Figure 1), the vector aiming for the flap pocket was varied systematically to explore the binding properties of this cavity, especially with respect to the volume that is available for binding.

In this context, important concepts previously demonstrated in molecular recognition studies with synthetic receptors were transferred and applied to an enzyme environment. The 55% rule of Mecozzi and Rebek states that inclusion complexes are favored, when a guest occupies 55±9% of the available space within a host.[11] This rule applies in particular to apolar binding processes.[12–19] In the case of alkyl chains, it has been observed that residues apparently too long in their fully extended conformation for a certain cavity, adopt energetically less favorable conformations to fit into the available space.[15–22] Unfavorable internal *gauche* strain in these contorted guests is compensated by burial of hydrophobic surfaces, chemical complementarity, and proper filling of space. Examples of nonstaggered alkyl residues bound by enzymes such as lipid-binding proteins are also known.[23,24] It was further suggested that the principle of ideal volume occupancy is generally applicable, for example, to drug design.[11] However, the 55% rule has to our knowledge not yet been applied to enzymes. In contrast to rigid synthetic container molecules, proteins are more flexible, and there are several cases known where the enzyme adapts itself to the—although contorted—ligand.[23] It is therefore difficult to estimate the volume of an enzyme subpocket. The flap pocket of PM II however is relatively narrow and well defined, as it is observed in the three X-ray crystal structures of PM II

[a] *M. Zürcher, T. Gottschalk, Prof. Dr. F. Diederich*
Laboratorium für Organische Chemie, ETH Zürich
Wolfgang-Pauli-Strasse 10, 8093 Zürich (Switzerland)
Fax: (+41) 44-632-1109
E-mail: diederich@org.chem.ethz.ch

[b] *Dr. S. Meyer, Dr. D. Bur*
Drug Discovery, Chemistry and Biology, Actelion Pharmaceuticals Ltd.
Hegenheimermattweg 91, 4123 Allschwil (Switzerland)

Supporting information for this article is available on the WWW under http://www.chemmedchem.org or from the author.

with inhibitors featuring *n*-pentyl substituents (PDB codes 2BJU, 2IGX, and 2IGY, see the Supporting Information for an overlay).[7,8]

We used these three structures to determine the cavity volume of the flap pocket by two different computational methods: a) filling the cavity with a hydrocarbon network[11] and b) an automated procedure implemented in the software HOLE[25] (for details see the Supporting Information). Both approaches gave comparable results, and a mean value of 252 ± 6 Å^3 (error given as standard deviation) was obtained for the three structures 2BJU, 2IGX, and 2IGY. The three calculated values are very similar (see the Supporting Information), which confirms that the positions of the amino acids defining the volume of the flap pocket are indeed conserved throughout these three structures. In the modeling, the first methylene group of the alkyl chain flap vector is surrounded by Trp 41, Met 75, Asn 76, Tyr 77, Phe 111, and Ile 123, and the plane through these amino acid residues was therefore assigned to delineate the flap pocket of PM II (see Supporting Information Figure 3SI).

Modeling with MOLOC[26] suggests that no fully extended *n*-alkyl chain flap vector longer than seven C atoms fits into the flap pocket of PM II. The distances of the *n*-heptyl chain of inhibitor (±)-**3** to the "back end" of the cavity imply that longer chains have to fold in order to be accommodated in the pocket (Figure 2). Such a distortion is indeed predicted for the longer chains in the molecular modeling (Figure 3).

Compounds (±)-**3**–(±)-**14** with different (cyclo)alkyl substituted alkyl chain flap vectors were synthesized[27] and tested in vitro for inhibition of PM II and IV (Table 1). For compounds (±)-**10** and (±)-**12**–(±)-**14**, a clear trend can be seen for both plasmepsins: the bigger the substituent at the vector end, the higher the IC_{50} values. This implies that there is only limited space at the "end" of the flap pocket, and that the cavity cannot easily reorganize to accommodate the more voluminous substituents. However, if the chain is shortened, a cyclopentyl ring fits well in the cavity (compare (±)-**11** and (±)-**13**).

A striking structure–activity relationship becomes evident for a series of inhibitors (±)-**1** and (±)-**3**–(±)-**9**, which target the flap pocket with a homologous series of *n*-alkyl chains. Plotting the binding affinities (IC_{50} values) for both enzymes against the chain length of the flap vector results in bell-shaped curves with broad maxima (Figure 4).

For PM II, the affinity initially increases with the number of C atoms (*n*) in the chain, until ≈50% of the cavity volume is filled in the case of (±)-**3** (R^{flap} = *n*-heptyl). Ligand (±)-**3** shows the strongest affinity in this series (IC_{50} = 50 nM), whereas (±)-**6** (R^{flap} = *n*-octyl) and (±)-**7** (R^{flap} = *n*-nonyl) are also bound with high strength. Their alkyl chains do still fit well inside the cavity (packing coefficient (PC) = 0.57 and 0.63, respectively), and the IC_{50} values are 70 and 150 nM. According to modeling studies, the longer alkyl chains of (±)-**6** and (±)-**7**, in contrast to (±)-**3**, have to reduce their length by folding. Binding rapidly decreases if the chain length exceeds nine methylene groups: inhibitors (±)-**8** (R^{flap} = *n*-decyl) and (±)-**9** (R^{flap} = *n*-undecyl) are bound with only moderate affinities (IC_{50} = 350 and 1950 nM, respectively), because the *n*-alkyl chains of (±)-**8** and

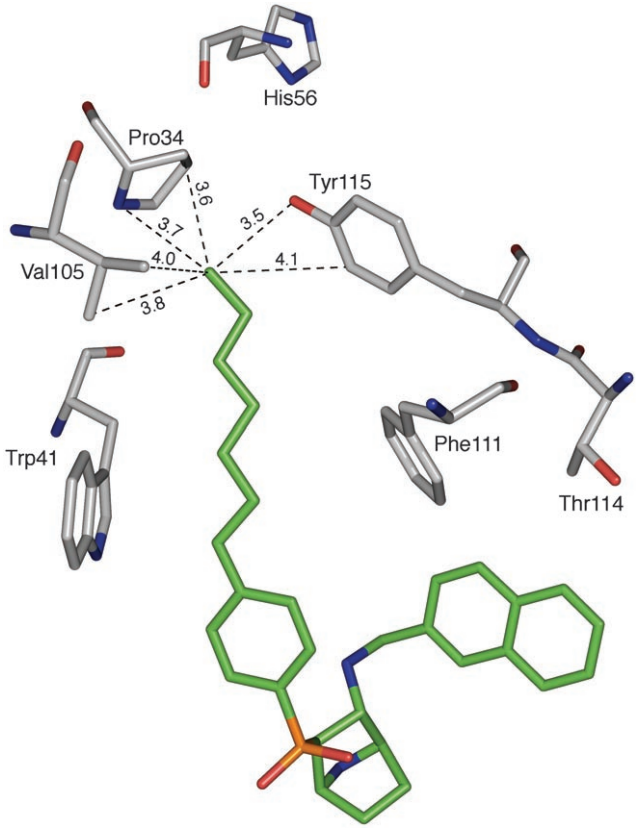

Figure 2. MOLOC-generated molecular model of inhibitor (±)-**3** showing important amino acids surrounding the flap pocket (PDB code 2BJU). Color code: C-skeleton of enzyme: gray, C-skeleton of (±)-**3**: green, O-atoms: red, N-atoms: blue, S-atom: orange. The enantiomer of the inhibitor that is predicted by modeling to bind into the active site is shown. Distances are indicated in Å.

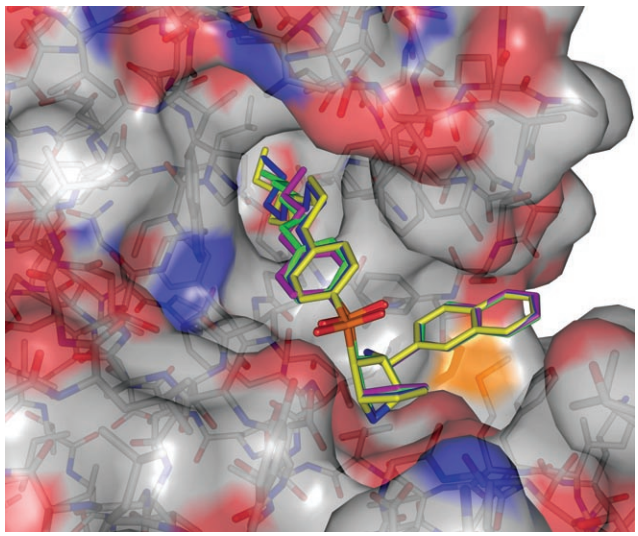

Figure 3. MOLOC-generated molecular model of inhibitors featuring *n*-alkyl chain flap vectors bound within the active site of PM II (PDB code 2BJU). Several amino acid residues have been removed to permit better visualization of the flap pocket. Color code: C-skeleton of enzyme: gray, C-skeleton of (±)-**4**: purple, C-skeleton of (±)-**3**: green, C-skeleton of (±)-**7**: blue, C-skeleton of (±)-**8**: yellow, O-atoms: red, N-atoms: blue, S-atoms: orange. The enantiomer of the inhibitor that is predicted by modeling to bind into the active site is shown.

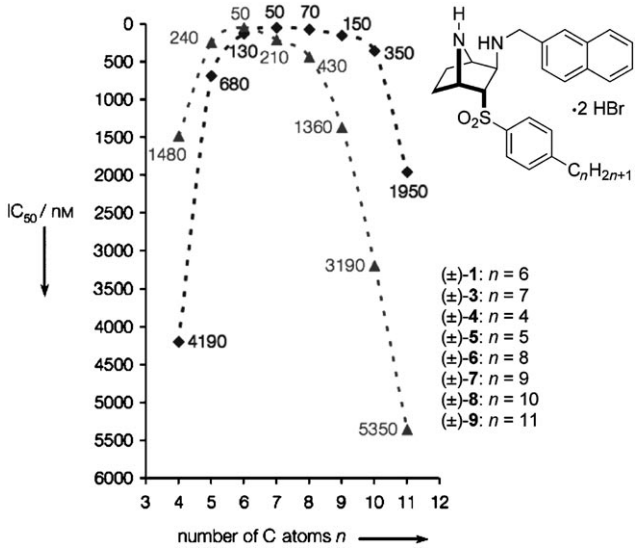

Figure 4. IC_{50} values [nM] for PM II (squares) and IV (triangles) as a function of the number of C atoms *n* in the *n*-alkyl chain flap vector.

(±)-**9** are too large to be accommodated by the flap pocket even in their folded conformations (PC = 0.68 and 0.75, respectively).

The hypothesis that *n*-alkyl chains adopt their conformation to fill the flap pocket ideally is corroborated by the fact that the calculated PCs lie in the favorable range of 55 ± 9% for our best PM II inhibitors. Molecules, which appear to be too long to fit inside the flap pocket, are among the best binders if their volume occupation is in the preferred range. Taking these arguments into account helps to explain the rather broad curve maximum spanning six differently long chains (PM II, Figure 4).

The ideal chain length is different for PM II and PM IV respectively. Not only does PM IV prefer shorter chains, the decrease in activity when exceeding optimal chain length is also much stronger for PM IV than for PM II (Figure 4 and Table 1). As a result of the lack of X-ray crystal structural data, volume considerations for the flap pocket could not be made in the case of PM IV. Based on the biological results however, PM IV seems to have a shorter flap pocket with a smaller binding volume if compared to PM II.

As a note of caution, the volume of an enzyme pocket has to be taken as a good estimation rather than a measurement and is therefore less reliable than the respective values determined for rigid synthetic hosts. A well-defined and highly conserved binding pocket is needed for exact determinations.

In conclusion, different inhibitor chain lengths yielded optimal results for PM II and IV respectively, which suggests significant differences in the size and shape of their flap pockets. The 55% rule of Mecozzi and Rebek has been successfully applied to an enzyme environment. Similar to synthetic host–guest binding events, long *n*-alkyl groups in the flap pocket of PM II seem to be readily contorted, if this is required to fit inside the binding site, and if volume occupation in the resulting complex is in the preferred range. Volume considerations can provide valuable information whether a lipophilic enzyme or receptor pocket is optimally filled, and we foresee growing importance of volume analyses as a tool in structure-based design.

Acknowledgements

M.Z. thanks the Roche Research Foundation for a Ph.D. fellowship. This work was supported by the ETH Research Council.

Keywords: aspartic proteases · drug design · inhibitors · malaria · molecular recognition

[1] WHO fact sheet N° 94, **2007**; available online: http://www.rollbackmalaria.org/cmc_upload/0/000/015/372/RBMInfosheet_1.htm.
[2] R. W. Snow, C. A. Guerra, A. M. Noor, H. Y. Myint, S. I. Hay, *Nature* **2005**, *434*, 214–217.
[3] J. Liu, I. Y. Gluzman, M. E. Drew, D. E. Goldberg, *J. Biol. Chem.* **2005**, *280*, 1432–1437.

Table 1. Inhibition of PM II and IV and calculated PCs for PM II.

	Nr	R^{flap} =	PC [%] PM II	IC_{50} [nM] PM II[a]	IC_{50} [nM] PM IV[a]
	(±)-**4**		30	4190	1480
	(±)-**5**		36	680	240
	(±)-**1**		43	130	50
	(±)-**3**		50	50	210
	(±)-**10**		50	30	90
	(±)-**11**		52	30	50
	(±)-**12**		53	170	180
	(±)-**6**		57	70	430
	(±)-**13**		59	190	210
	(±)-**7**		63	150	1360
	(±)-**14**		65	380	760
	(±)-**8**		68	350	3190
	(±)-**9**		75	1950	5350

[a] Values represent the average of two repetitions of a fluorogenic proteolysis assay; the estimated error for this assay is ±50%. PCs were calculated using optimized inhibitor structures (see the Supporting Information).

[4] A. L. Omara-Opyene, P. A. Moura, C. R. Sulsona, J. A. Bonilla, C. A. Yowell, H. Fujioka, D. A. Fidock, J. B. Dame, *J. Biol. Chem.* **2004**, *279*, 54088 – 54096.
[5] O. A. Asojo, S. V. Gulnik, E. Afonina, B. Yu, J. A. Ellman, T. S. Haque, A. M. Silva, *J. Mol. Biol.* **2003**, *327*, 173 – 181.
[6] J. C. Clemente, L. Govindasamy, A. Madabushi, S. Z. Fisher, R. E. Moose, C. A. Yowell, K. Hidaka, T. Kimura, Y. Hayashi, Y. Kiso, M. Agbandje-McKenna, J. B. Dame, B. M. Dunn, R. McKenna, *Acta Crystallogr., Sect. D.* **2006**, *62*, 246 – 252.
[7] L. Prade, A. F. Jones, C. Boss, S. Richard-Bildstein, S. Meyer, C. Binkert, D. Bur, *J. Biol. Chem.* **2005**, *280*, 23837 – 23843.
[8] C. Boss, O. Corminboeuf, C. Grisostomi, S. Meyer, A. F. Jones, L. Prade, C. Binkert, W. Fischli, T. Weller, D. Bur, *ChemMedChem* **2006**, *1*, 1341 – 1345.
[9] C. Oefner, A. Binggeli, V. Breu, D. Bur, J.-P. Clozel, A. D'Arcy, A. Dorn, W. Fischli, F. Grüninger, R. Güller, G. Hirth, H. P. Märki, S. Mathews, M. Müller, R. G. Ridley, H. Stadler, E. Vieira, M. Wilhelm, F. K. Winkler, W. Wostl, *Chem. Biol.* **1999**, *6*, 127 – 131.
[10] F. Hof, A. Schütz, C. Fäh, S. Meyer, D. Bur, J. Liu, D. E. Goldberg, F. Diederich, *Angew. Chem.* **2006**, *118*, 2193 – 2196; *Angew. Chem. Int. Ed.* **2006**, *45*, 2138 – 2141.
[11] S. Mecozzi, J. Rebek, Jr., *Chem. Eur. J.* **1998**, *4*, 1016 – 1022.
[12] L. Pirondini, D. Bonifazi, B. Cantadori, P. Braiuca, M. Campagnolo, R. De Zorzi, S. Geremia, F. Diederich, E. Dalcanale, *Tetrahedron* **2006**, *62*, 2008 – 2015.
[13] T. Gottschalk, B. Jaun, F. Diederich, *Angew. Chem.* **2007**, *119*, 264 – 268; *Angew. Chem. Int. Ed.* **2007**, *46*, 260 – 264.
[14] O. Hayashida, L. Sebo, J. Rebek, Jr., *J. Org. Chem.* **2002**, *67*, 8291 – 8298.
[15] A. Scarso, L. Trembleau, J. Rebek, Jr., *Angew. Chem.* **2003**, *115*, 5657 – 5660; *Angew. Chem. Int. Ed.* **2003**, *42*, 5499 – 5502.
[16] L. Trembleau, J. Rebek, Jr., *Science* **2003**, *301*, 1219 – 1220.
[17] A. Scarso, L. Trembleau, J. Rebek, Jr., *J. Am. Chem. Soc.* **2004**, *126*, 13512 – 13518.
[18] B. W. Purse, J. Rebek, Jr., *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 2530 – 2534.
[19] R. J. Hooley, J. Rebek, Jr., *Org. Lett.* **2007**, *9*, 1179 – 1182.
[20] D. Ajami, J. Rebek, Jr., *J. Am. Chem. Soc.* **2006**, *128*, 15038 – 15039.
[21] K. A. Udachin, G. D. Enright, E. B. Brouwer, J. A. Ripmeester, *J. Supramol. Chem.* **2001**, *1*, 97 – 100.
[22] O. Sénèque, M.-N. Rager, M. Giorgi, O. Reinaud, *J. Am. Chem. Soc.* **2000**, *122*, 6183 – 6189.
[23] M. P. Schramm, J. Rebek, Jr., *Chem. Eur. J.* **2006**, *12*, 5924 – 5933.
[24] J. Rebek, Jr., *Chem. Commun.* **2007**, 2777 – 2789.
[25] O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, M. S. P. Sansom, *J. Mol. Graphics* **1996**, *14*, 354 – 360.
[26] P. R. Gerber, K. Müller, *J. Comput.-Aided Mol. Des.* **1995**, *9*, 251 – 268.
[27] The synthesis of (±)-**3**–(±)-**14** follows the protocol described for (±)-**1** in [10]. All compounds were fully characterized by melting points, ^{1}H and ^{13}C NMR spectroscopy, IR, MS, and high-resolution MS and/or elemental analysis. A full account on synthesis and characterization will be part of a future full paper.

Received: September 3, 2007
Published online on October 4, 2007

