J Mol Model (2011) 17:769–775
DOI 10.1007/s00894-010-0732-6

ORIGINAL PAPER

Molecular dynamics investigation of psalmopeotoxin I. Probing the relationship between 3D structure, anti-malarial activity and thermal stability

Matthew Paul Gleeson · Songpon Deechongkit · Somsak Ruchirawat

Received: 3 March 2010 / Accepted: 23 April 2010 / Published online: 12 June 2010

Abstract PcFK1 is a member of the cysteine knot inhibitor family that displays anti-malarial properties. The naturally occurring molecule is ~40 amino acids in length and forms a highly constrained 3D structure due to the presence of 3 disulfide and multiple *intra*-molecular H-bonds. Recent experimental studies on PcFK1 wild-type and mutants, where the cystiene residues of each disulfide bond were mutated into serine residues, suggest that alterations to these structural constraints can give rise to sizeable differences in SAR. To better understand the relationship between the dynamic inhibitor 3D structure, biophysical and biological properties we have performed solution based molecular dynamics calculations over 150ns using the CHARMM forcefield. We have analyzed the theoretical trajectory in a systematic way using principal components analysis, which allows us to identify the correlated nature of the protein loop, turn and sheet movements. We have identified the key molecular motions that give rise to the differing SAR which has helped to more precisely direct our ongoing SAR studies in this important therapeutic area.

Keywords Anti-malaria · CHARMM · Molecular dynamics · PCA · PcFK1 · SAR · Tm

Introduction

Malaria is one of the most widespread infectious diseases in the developing world [1, 2]. While the disease mechanism is well understood and drugs molecules are available to treat the disease, the ever increasing resistance to these treatments means that more effective therapeutic agents are constantly being sought [3]. Hampering this effort is the fact that the infection is concentrated in poor, developing nations, particularly in Africa [4], meaning the pressure on industry to research and subsequently develop new medicines is relatively low. Fortunately, a significant amount of academic research is being done in this area to both develop greater understanding of the disease mechanism as well as novel molecular leads.

Psalmopeotoxin I (PcFK1) is a 33-amino-acid peptide which is isolated from the venom of *tarantula Psalmopoeus cambridgei* [5]. It is a member of the Inhibitor Cystine Knot (ICK) superfamily that all contain three conserved disulfide bridges. These are critical in defining their overall structural tertiary structure [6, 7]. PcFK1 and related analogs have shown interesting anti-malarial activity in the previous reports making them a possible candidate for the required future generation of treatments [5–9]. Unfortunately, the

Electronic supplementary material The online version of this article (doi:10.1007/s00894-010-0732-6) contains supplementary material, which is available to authorized users.

M. P. Gleeson (✉)
Department of Chemistry, Faculty of Science,
Kasetsart University,
50 Phaholyothin Rd, Chatuchak,
Bangkok 10900, Thailand
e-mail: paul.gleeson@ku.ac.th

S. Deechongkit · S. Ruchirawat
Laboratory of Medicinal Chemistry,
Chulabhorn Research Institute,
54 Moo 4, Vibhavadee-Rangsit Highway, Laksi,
Bangkok 10210, Thailand

S. Deechongkit · S. Ruchirawat
Chulabhorn Graduate Institute and the Center for Environmental Health, Toxicology & Management of Chemicals,
54 Moo 4, Vibhavadee-Rangsit Highway, Laksi,
Bangkok 10210, Thailand

molecular target of PcFK1 is not currently known, restricting the rational design of new analogs to ligand based design methods. However, computational ligand based structure-activity-relationships (SAR) for complex natural molecules, such as that for the ICK family, can be extremely challenging to derive due to the many degrees of freedom present. The issue is that the computed properties typically used to derive the relationship are based on 2D structure, or based on a single, static 3D structure, which can be a gross approximation for flexible, dynamic molecules [10]. In this case properties derived from molecular dynamics (MD) simulations may be preferable since these can capture the dynamic nature of the molecules in question.

In this paper we discuss the application of MD simulations to PcFK1 to help rationalize SAR data recently reported [9] (Table 1). This work describes alterations to the wild-type peptide sequences that give rise to both increases and decreases in anti-malaria activity, as well as changes to thermal stability (T_m). To better understand how the differences in structure affect the dynamic properties of the molecules, and in turn how this impacts on the biophysical/activity responses, we have performed theoretical simulations using the CHARMM program [11].

Computational procedures

The NMR structure of PcFK1 was downloaded from the RCSB protein databank [12] (accession code: 1X5V [8]). All calculations were performed on model_1. Disulfide bonds between sulfur atoms on residues 2&19, 9&24 and 18&29 were added using Discovery Studio 2.1 [13]. The structure was solvated in an octahedral box of water (n=1379, 7Å to edge of box) and neutralized by the addition of Na^+ (n=3) and Cl^-(n=5) counterions to give a net salt concentration of 0.145 M. Models for three additional analogs/mutants (Table 1) were constructed in a similar manner by mutating both cysteine residues associated with each disulfide bond in turn to serine.

Table 1 Physical and biological results obtained for the PcFK1 wild-type and mutant inhibitors. Data taken from reference 9. %A values taken at 20 μg/mL peptide concentration

PcFK1	Mutation	T_m (°C)	Activity (%)
Wild type	–	14.4	59.4
Mutant 1	C9S, C24S	6.2	62.7
Mutant 2	C18S, C29S	13.9	87.3
Mutant 3	C2S, C19S	7.9	65.4

Proteins were parameterized using the CHARMM22 [14] forcefield and solvent using the TIP3P method [15]. Models were initially optimized to an RMS gradient below 0.1 kcal mol^{-1}. MD was performed in three stages; (a) heating from T=0 to 300 K over 2 nS (b) equilibration for 50nS and (c), production dynamic step for 100nS. All simulations were performed using the default CHARMM settings. These include a time step of 0.001 pS, NVT conditions, 12Å non-bonded cut-offs, particle mesh Ewald [16] and shake constraints on hydrogen atoms. Structures were written every 0.1nS during the production run for further analysis.

The *intra*-molecular H-bonds and disulfide bonds formed by the four different PcFK1 structures determine the conformation of this tightly compacted peptide inhibitor. To understand how the structure changes over time, and the correlated nature of the movements of the different structural elements within the molecule, we have extracted all backbone H-bond interactions (*i.e.*, β-sheets/turns) as well as the disulfide bonds distances from each model (Fig. 1). In certain cases a new interaction forms in only one mutant structure and in this case we also extract the corresponding distance in those in which it does not. We focus on the *intra*-molecular H-bond distances rather than atomic RMSDs since they contain a greater amount of information in a smaller number of variables (essentially a vector). The correlated nature of the changes in the interactions over the course of the simulation were then analyzed using PCA [17, 18]. A model was built using the averaged parameter data in SPSS16.0 [19]. Input descriptors were mean centered and scaled. Models were built using the default settings.

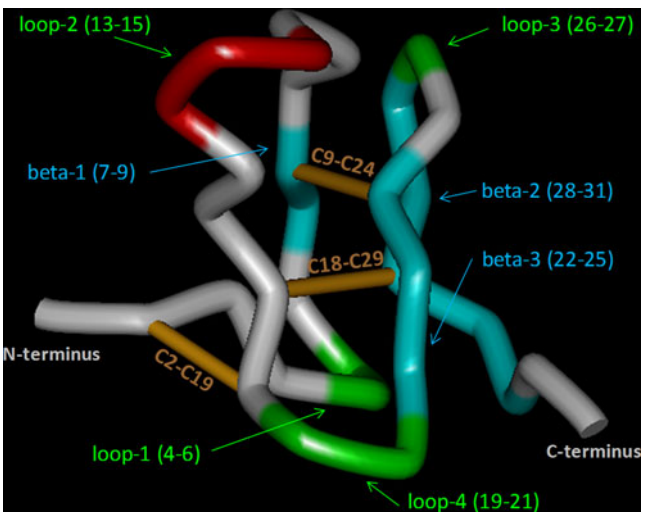

Fig. 1 A simplified representation of the key features of PcFK1. Diagram is colored according to secondary structure elements: β-sheets, cyan, β-turns, green/red, and disulfide bridges, brown

Results and discussion

The solution NMR structure of PcFK1 is similar to other ICK family peptides in that it is a densely packed peptide displaying a number of β-sheet and β-turns features. These are formed as a result of 3 disulfide bonds and a number of *intra*-molecular H-bond interactions. Due to its small size, the hydrophobic core is relatively small compared to other proteins. We have captured the dynamic nature of these molecules over the course of a 150 nS MD simulation. The results are summarized in graphical form in Fig. 2 as well as in the supplementary information (Fig. S1, total energy and RMSD). The average parameters associated with the key interactions over the 100 nS of MD are also summarized in Table 2.

Molecular flexibility

The inhibitors show sizeable differences in their *intra*-molecular interactions and these are not simply limited to regions where the disulfide bonds have been broken. The results suggest that each mutation can have a dramatic effect on the overall tertiary structure, but that the location of the change also affects the structure in subtly different, but often correlated ways.

In Fig. 2a a subset of the *intra*-molecular interactions are highlighted for the wild-type (WT). The interaction distances remain essentially constant over the course of the simulation suggesting it is a quite stable structure in solution. However, breaking the Cys9-Cys24 disulfide bond (through serine mutation) leads to a sizeable change in the *intra*-molecular parameters (mutant-1, Fig. 2b). It also leads to an increase in the overall number of H-bonds made by the inhibitors compared to the WT as a result of greater solvent exposure (Table 2). When compared to the WT, the average distance between Ser9-Ser24 in mutant-1 increases by ~0.9 Å as does the interaction between Arg25 and Lys28 found at the top of loop-3/*β*-hairpin (Table 2). We also see decreased intermolecular interactions (increasing distance) between Val12-Asn16 and Val10-Val12, which are located in the region termed loop-2 (Fig. 1). In contrast, the interaction between Ile4-Asp7 in loop-1 increases somewhat (Table 2). The mutation does not appear to change the *intra*-molecular interactions formed

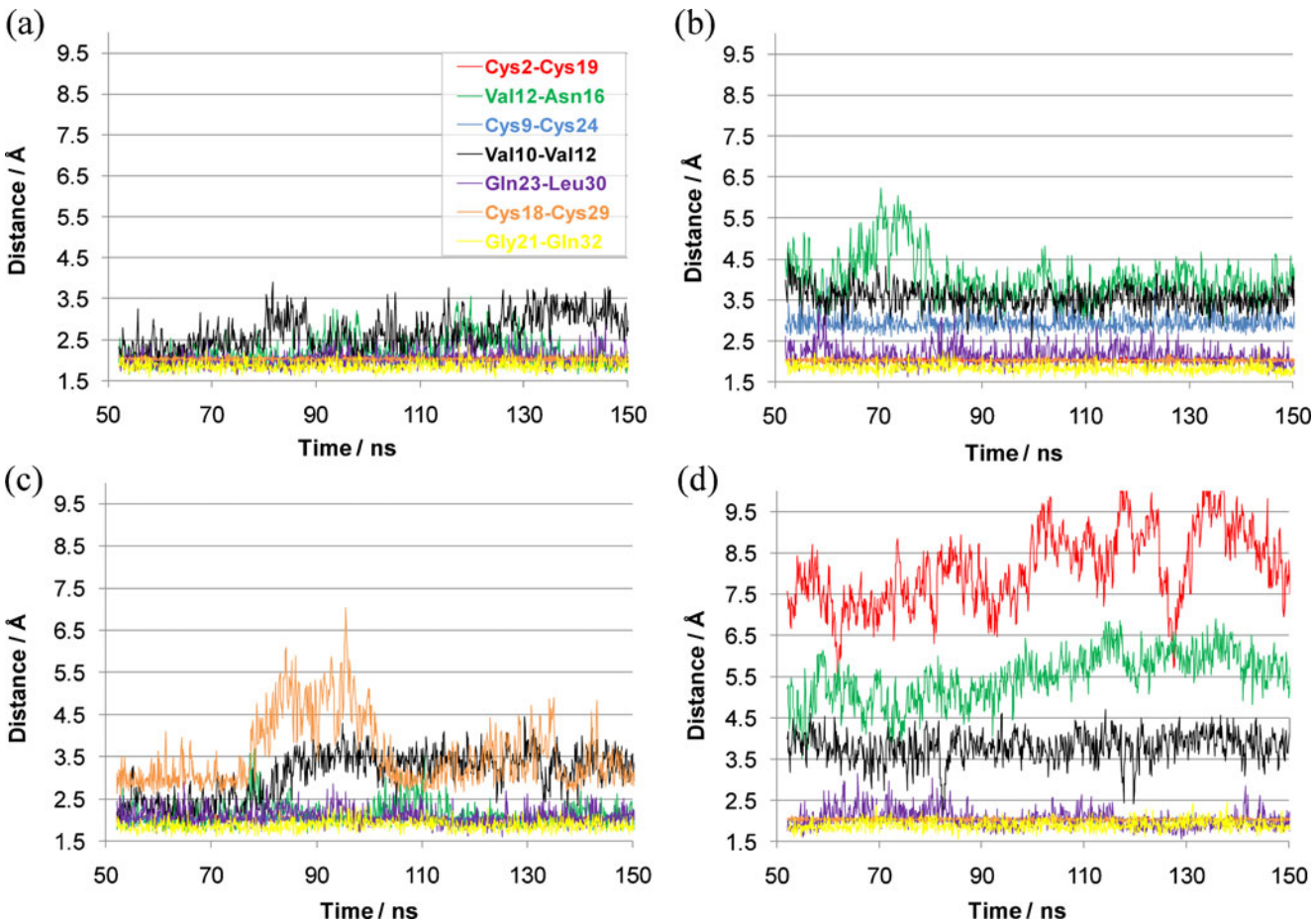

Fig. 2 The most variable PcFK1 interactions observed over the 100 nS of production dynamics. (**a**) corresponds to the WT, (**b**) Mut-1, (**c**) Mut-2 and (**d**) Mut-3

Table 2 Mean H-bond interaction and disulfide bond distances (Å) observed over the 100 nS of production dynamics. For mutants, the Cys-Cys value given refers to the distance between serine O atoms

ID	WT	Mut-1	Mut-2	Mut-3
Cys2-Cys19[1]	2.03	2.02	2.02	8.19*
Gly3-Pro17	1.90	2.03	1.87	2.01
Gly3-Cys19	2.09	2.13	2.77	4.75
Ile4-Asp7	2.52	2.24	2.15	2.69
Cys9-Cys24[1]	2.02	2.94	2.02	2.02
Cys9-Gly27	1.99	2.09	2.63	2.31
Val10-Val12	2.71	3.53	3.13	3.78
Val12-Asn16	2.21	4.01	2.14	5.52
Cys18-Cys29[1]	2.02	2.03	3.59*	2.03
Cys19-Leu22	2.27	2.22*	2.28	2.86
Gly21-Gln32	2.04	2.08	1.98	1.99
Gln23-Leu30 (1)	1.86	1.83	1.87	1.89
Gln23-Leu30 (2)	2.02	2.13	2.08	2.05
Arg25-Lys28 (1)	2.28	2.2	1.87	2.35
Arg25-Lys28 (2)	2.11	2.80	2.59	2.80
Cα RMSD	2.57	2.70	2.79	2.50
Total H-bonds	12.9	14.9	15.1	11.6

by the *beta*-sheet (between units a&c) and β-hairpin (b&c) to a significant degree.

In mutant-2, the disulfide bond (Cys18-Cys29) that underwent mutation lies in a more central location in the molecule. In this case one observes a much larger increase in distance between the two residues involved compared to mutant-1 (3.59 *vs*. 2.94 Å respectively). In mutant-2 we also see a decrease in the interaction toward the end of the β-hairpin (Gly3-Cys19), and a corresponding increase in interaction between residues of loop-1 (Ile4-Asp7) (Table 2). Another observation is the decrease in interaction at the top of the β-sheet (Cys9-Gly27). This change is perhaps surprising given it proximity to the disulfide bond between Cys9 and Cys24 showing the subtle affect of a relatively distant structural changes.

The disulfide bond removed in mutant-3 links the N-terminus of the peptide inhibitor to loop-4 (Cys2-Cys19). It is not involved in the formation of any major secondary structural elements which is surprising given the dramatic changes in tertiary structure observed over the course of the MD simulations. The loss of the disulfide bond actually leads to a dramatic increase in flexibility of residues in the N-terminal region and loop-4 up to the disulfide bond formed at residue 9. The mutation leads to the loss of structure in loop-1 (*i.e*. Gly3-Gly7 mean distance=4.75 Å) and the loss of interaction between the N-terminal domain/loop-1 and loop-4 (Cys2-Cys9 mean distance =8.2 Å). The interaction distance between residues in the *beta*-hair pin (Arg25-Lys28) increase by 0.8 Å over the WT structure as do that of Cys9-Gly27 in the β-sheet (1.99 to 2.31 Å respectively).

Molecular dynamics SAR

Kamolkijkarn *et al*. [9] reported that the anti-malarial activity and thermal melting SAR of the PcFK1 inhibitors under investigation were not correlated, suggesting they have a different underlying relationship with molecular structure. In an attempt to rationalize these SAR differences we reanalyzed the activity and biophysical data in the context of our MD simulation using PCA. PCA is a popular unsupervised statistical method used in a wide variety of informatics applications. In this case we can use the method to assess the correlated nature of the structural changes observed for the inhibitors under investigation here. We can use PCA to determine whether (a) one part of the molecule flexes in conjunction with a bend or compression elsewhere and (b) how correlated are the motions between the wild-type and mutants. This structural clustering may then help us to understand the SAR differences

The PCA model generated from the MD derived distance parameters uses two components to explain 91% of the total variation in the dataset (82% & 9% for components 1 and 2 respectively). Briefly, PCA components are linear combinations of the original descriptors. Components are comprised of correlated descriptors and are extracted in order of the total amount of information they explain. Each component extracted is orthogonal to each other (*i.e*., identifies a different underlying signal). The similarity between the different inhibitors can be determined from the PCA scores plot (Fig. 3a&b). The loadings plot describes the correlation between the different descriptors. (Fig. 3c&d). Observations found on the scores plot, or descriptors on the loading plot, that are located close to each other in both dimensions are highly correlated. Those that are similar on just one component are still correlated, but to a lesser extent. Those at opposite ends of a component are inversely correlated. Furthermore, descriptors found at the extreme ends of the X or Y axes have the most significant impact on the component that is defined by that axis. Those found close to the origin have a negligible effect. Finally, observations in the scores plot with a positive value on component 1 are positively correlated with descriptors on the loading plot with a positive value on component 1 (and vice versa).

The PCA results show that the majority of the variation in the 2D H-bond and disulfide bond distances can be described by two orthogonal components. We can in fact associate these two orthogonal components with two distinct 3D motions, offering a simple graphical way to view the structural differences. Component one essentially captures the changes in structure corresponding to the

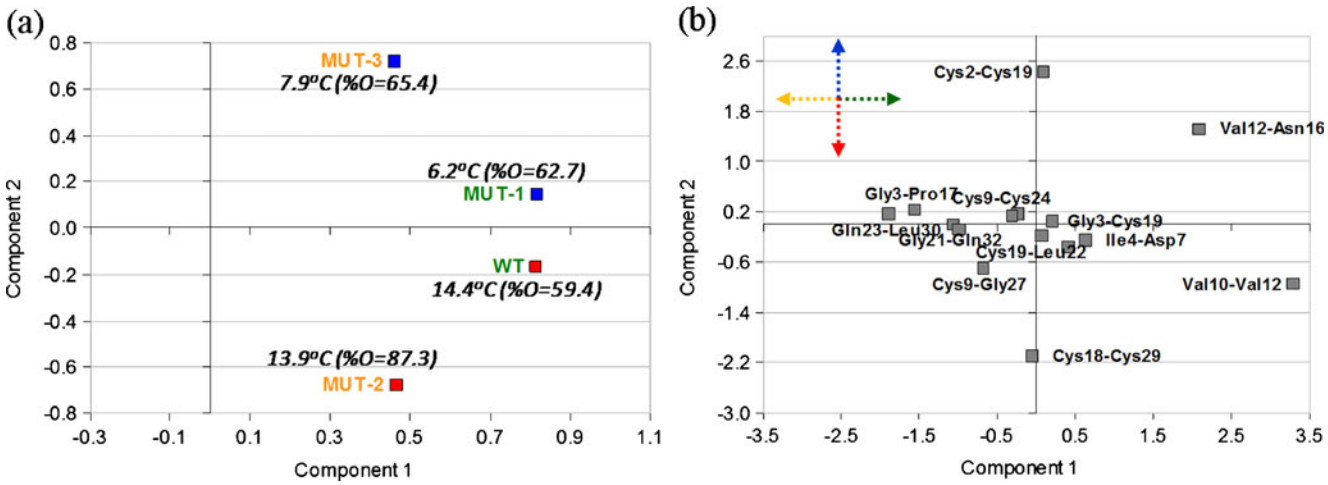

Fig. 3 PCA scores (**a**) and loading plots (**b**) obtained from an analysis of the MD derived structural parameters. Component 1 describes 91% and component 2, 8% of the total variance in the dataset

elongation of the molecule perpendicular to the disulfide bonds (*i.e.*, descriptors and observation with+ve loadings in Fig. 3b, corresponding to movements in the Z-axis as defined in Fig. 4a). This motion is the predominant dynamic change observed in these structures over the course of the MD simulation. Mutant-2 and mutant-3 have less positive loadings than the wild-type and mutant-1 on component one. This means that the wild-type and mutant-1 are more compressed in the Z direction than the other mutants (*i.e.* perpendicular to the disulfide bonds). Analysis of the distances in Fig. 3b reveals that as β-hairpin opens, there is a decrease in the interaction between loop1 and loop4, and a corresponding increase in interaction between loop-2 intermolecular interactions on average.

Component two is a minor motion compared to component one, which might be expected since it is essentially parallel to the much stronger disulfide bonds (movement along the X-axis as defined in Fig. 4b). The structural characteristics identified by component two primarily describe the major differences separating mutant-2 and mutant-3. Cys2-Cys19 and Cys18-Cys29 distances are found to be orthogonal to each other on component two. This is due to the fact that deletion of the Cys2-Cys19 bond leads to a larger decrease in the interaction between Val12-Asn16 in loop2 and Val10-Val12 in loop1, when compared to Cys18-Cys29 bond deletion in mutant-2.

From the PCA results, it is apparent there are two major concerted motions observed in these highly constrained peptide inhibitors over the course of the MD simulations.

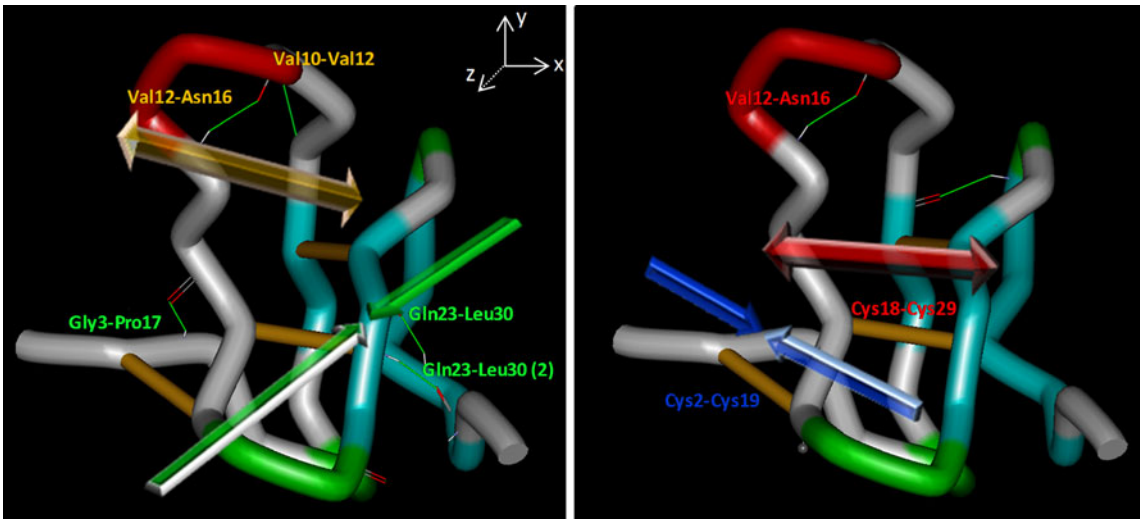

Fig. 4 The two distinct dynamics motions associated with PCKF1 as summarized from the PCA analysis. For component 1 the residues/distances in orange expand while those in green contract. For component 2, the residues/distances in blue contract while those in red expand

The two orthogonal components identified by PCA from the 2D descriptors in fact correspond to the two essentially orthogonal 3D motions of the proteins. These motions are a result of the unique arrangement of their structural elements, which leads to concerted molecular movements orthogonal to the disulfide bonds and perpendicular to them, with minimal movement in the 3rd dimension. To understand whether these structural characteristics could help to rationalize the SAR differences, the thermal melting and anti-malaria activity data were superimposed on the PCA scores plot (Fig. 3a). Mutant-1 and WT are separated from the more weakly potent mutant-2 and mutant-3 on component one, suggesting that inhibitors that can more effectively suppress motion perpendicular to the disulfide bonds will have better activity. In contrast, mutant-2 and mutant-3 are separated on component three from the considerably more stable mutant-2 and the WT. This suggests that greater flexibility in the direction parallel to the disulfide-bond will lead to lower thermal melting temperatures.

The qualitative agreement observed between the experimental activity and thermal melting data and the theoretical MD results can be rationalized in a relatively straightforward manner. Deletion of a disulfide bond will increase the flexibility of these peptide inhibitors parallel to the broken bonds (*i.e.*, component two). Since these bonds are critical for maintaining the constrained structure of these molecules, it is not surprising therefore that the inhibitors showing the greatest structural deviation parallel to the disulfide bonds over the MD simulation generally have the lowest thermal melting temperature. Furthermore, given that the thermal melting temperature does not correlate with the anti-malaria activity, and that there are only two significant motions in the molecules, then it might be logical to assume that this 2nd motion might explain the latter SAR differences. This is in fact what we observe. We find that molecules with lower flexibility perpendicular to the disulfide bonds (*i.e.*, component one) are generally the more potent.

Conclusions

This theoretical study has focused on a naturally derived spider toxin and related derivatives which display useful anti-malarial activity. We have performed MD simulations of the wild-type and three mutants to understand their dynamic properties and how these relate to both biophysical and biological properties. We subsequently extracted the key intermolecular H-bond and dis-sulfide bond distances over the 100nS MD simulation to gain insight into the peptide inhibitor intrinsic flexibility. Principal components analysis was used to assess the molecular motions observed over the course of the MD simulation revealing two distinct orthogonal molecular motions, one parallel to the disulfide bonds and one perpendicular.

The results of this study suggest that the different mutations have a dramatic effect on the overall tertiary structure, but that the location of the change also affects the structure in subtly different, but often correlated ways. It appears that minimizing flexibility perpendicular to the direction of the disulfide bonds is important to maintain higher thermal stabilities, while minimizing flexibility in the perpendicular direction appears to be important for good anti-malarial activity.

These results have helped us to gain insight into the physical basis of the thermal stability and activities of these interesting peptide inhibitors. While the SAR results may only be qualitative in nature, the development of a dynamic 3D model for such challenging molecules nevertheless gives us a way to theoretically assess new alternative mutants. This model provides us with a means to design mutants to target better thermal stability and anti-malarial activity, or design mutants that will have dramatically different dynamic properties to probe SAR space more widely, while synthesizing the fewest possible peptide mutants.

Acknowledgments We would like to acknowledge the support of NECTEC and NANOTEC who provided access to the Accelrys Software and Kasetsart University for the use of computational facilities. We would also like to acknowledge the support provided by the Chulabhorn Foundation and The Thailand Research Fund (RSA5180007).

References

1. Olliaro PL, Boland PB (2001) Antimalarial drug discovery: Old and new approaches. J Expt Biol 206:65–83. doi:10.1242/jeb.00589
2. Mendis K, Sina BJ, Marchesini P, Carter R (2001) The neglected burden of Plasmodium vivax malaria. Am J Trop Med Hyg 64:97–106
3. Ridley RG (2002) Medical need scientific opportunity and the drive for antimalarial drugs. Nature 415:686–693. doi:10.1038/415686a
4. Snow RW, Trape JF, Marsh K (2001) The past present and future of childhood malaria mortality in Africa. Trends Parasitol 17:593–597. doi:10.1016/S1471-4922(01)02031-1
5. Choi SJ, Parent R, Guillaume C, Deregnaucourt C, Delarbre C, Ojcius DM, Montagne JJ, Celerier ML, Phelipot A, Amiche M, Molgo J, Camadro JM, Guette C (2004) Isolation and characterization of Psalmopeotoxin I and II: Two novel antimalarial peptides from the venom of the tarantula Psalmopoeus cambridgei. FEBS Lett 572:109–117. doi:10.1016/j.febslet.2004.07.019
6. Corzoa G, Escoubas P (2003) Pharmacologically active spider peptide toxins. Cell Mol Life Sci 60:2409–2426. doi:10.1007/s00018-003-3108-6
7. Pallaghy PK, Nielsen KJ, Craik DJ, Norton RS (1994) A common structural motif incorporating a cystine knot and a triple-stranded

β-sheet in toxic and inhibitory polypeptides. Protein Sci 3:1833–1836
8. Pimentel C, Choi SJ, Chagot B, Guette C, Camadro JM, Darbon H (2006) Solution structure of PcFK1 a spider peptide active against Plasmodium falciparum. Protein Sci 15:628–634. doi:10.1110/ps.051860606
9. Kamolkijkarn P, Prasertdee T, Netirojjanakul C, Sarnpitak P, Ruchirawat S (2010) Deechongkit Synthesis, biophysical, and biological studies of wild-type and mutant psalmopeotoxins–anti-malarial cysteine knot peptides from Psalmopoeus cambridgei. Peptides. 31(4):533-540. doi:10.1016/j.peptides.2010.01.00
10. Hopfinger AJ, Wang S, Tokarski JS, Jin B, Albuquerque M, Madhav PJ, Duraiswami C (1997) Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism. J Am Chem Soc 119:10509–10524. doi:10.1021/ja9718937
11. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: A program for macromolecular energy minimization and dynamics calculations. J Comput Chem 4:187–217. doi:10.1002/jcc.540040211
12. RCSB Protein databank: http://www.rcsb.org/
13. Discovery Studio 2.1 Accelrys Inc, 10188 Telesis Court, Suite 100, San Diego, CA 92121, USA. www.accelrys.com
14. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616. doi:10.1021/jp973084f
15. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935. doi:10.1063/1.445869
16. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092. doi:10.1063/1.464397
17. Jackson JE (1991) A user's guide to principal components. Wiley, New York
18. Wold S, Geladi P, Esbensen K, Öhman J (1987) Multi-way principal components-and PLS-analysis. J Chemometr 1:41–46. doi:10.1002/cem.1180010107
19. SPSS 16.0, SPSS Inc, 233 S Wacker Drive, 11th floor, Chicago IL, 60606-6307, USA. www.spss.com