Yan and Wu, 2:3
http://dx.doi.org/10.4172/scientificreports.663

Open Access Scientific Reports

Research Article Open Access

2-Dimensional HP Folding Structures of Coenzyme Pyrroloquinoline Quinone Synthesis Protein A

Shaomin Yan and Guang Wu*

State Key Laboratory of Non-food Biomass Enzyme Technology, National Engineering Research Center for Non-food Biorefinery, Guangxi Key Laboratory of Biorefinery, Guangxi Academy of Sciences, 98 Daling Road, Nanning, Guangxi, 530007, China

Abstract

Pyrroloquinoline Quinone (PQQ) is a prosthetic group exists in bacterial methanol, ethanol and glucose dehydrogenases, and coenzyme PQQ synthesis protein A (PQQA) is a 24-amino acid peptide. Because of position of ethanol in bio-fuel industry, all these enzymes began to draw more and more attention on their structures and functions, especially folding structure. The hydrophobic-polar (HP) model is a simple model to study folding structure in 2- and 3-demnesional lattices. From a practical viewpoint, it would be interesting to study 2D folding structure of PQQA because 2D folding structure does not exist in nature, but sheds light on folding process. Moreover, HP model is computationally very challengeable because it belongs to NP problem. In this study, nine PQQAs and two variants were studied using 2D HP model with help of the normalized amino acid hydrophobicity index to convert 44 HP sequences, and each HP has 94,143,178,827 possible folding structures. The results show (i) PQQA has many native states with the same minimal energy, which may adopt for different enzymatic reactions; (ii) various symmetric folding structures exist, which help PQQA find its native state through different pathways; and (iii) the normalized amino acid hydrophobicity index can help furthermore distinguish native states numerically, which help to design native states with more negative energy. This study implies possible ways to modify enzymes through engineering.

Keywords: Folding structure; HP model; Hydrophobicity index; Minimal energy; Native state

Introduction

It is a consensus that bio-fuel should play a more and more important role in so-called green economy. In this context, all the enzymes either being related to ethanol itself or being related to any products connecting with ethanol could be interesting objectives for studies. A prosthetic group of bacterial methanol, ethanol and glucose dehydrogenases is Pyrrolo Quinoline Quinone (PQQ), whose trianionic form comes from deprotonation of three carboxy groups [1]. Although PPQ is a prosthetic group, there are many enzymes existing in bacteria for its synthesis. Of those enzymes, coenzyme PQQ synthesis protein A (PQQA) is a 24-amino acid peptide [2].

Such a small size peptide provides a good opportunity to apply various models to study its folding structures although it is easy to determine its 3-dimensional (3D) structure through experiments. Especially important is to study its 2-dimensional (2D) structures because 2D structure only exists in theoretical space, but its folding process is very suggestive to 3D structure and enzyme function.

Of theoretical models, the hydrophobic-polar (HP) model is a simple model, but is very difficult to apply to real-life cases. HP model was based on the observation that hydrophobic interaction was the driving force for protein folding and the hydrophobicity of amino acids was the main force to develop a native structure of small globular proteins [3,4].

Nevertheless, HP model has weaknesses, for example, its assumption appears simple, its description appears away from real-life case, etc. However, any model is an approximation to the real-life case, and difference between models essentially relies on the extent of their approximations. Naturally, if a model can provide insight, which cannot be obtained from other models, this model should have advantage.

Actually, HP model has yet to be fully studied because HP model needs extremely intensive computations, which was classified as NP problem [5,6], being the first problem listed millennium prize [7]. Hence, studies on HP model can help us understand NP problem, which leads to develop optimal algorithms [8-12] and deals with other intensively computational problems in biological fields such as phylogenetics, RNA pseudoknot [13].

HP model is workable for both 2D and 3D folding structures; however the number of computations for 3D is beyond our imagination. In 2D and 3D folding structures, each amino acid, either hydrophobic (H) or polar (P), walks along a line in 2D lattice or in 3D cube by taking a self-avoided step, then an H-H connection, which does not come from sequential step, is considered to have minus unity energy, and the structure with minimal energy is considered as a native folding structure.

In theory, the number of structures is 3^{n-1}, where n is the number of amino acid in a protein, for example, the number of structures for two amino acids is 3 (3^{2-1}), for three amino acids is 9 (3^{3-1}), for four amino acids is 81 (3^{4-1}),... for 24 amino acids is 94,143,178,827 (3^{24-1}). The Lenovo Think Pat laptop with due CPU of 2 GHz computed 200,000 to 250,000 folding structures per second, for PQQA with 24 amino acids, the computing time was between 94,143,178,827/250,000 and 94,143,178,827/200,000 seconds, which resulted in between 376,573 and 470,716 seconds, i.e., between 4.35 and 5.45 days. It is considered

***Corresponding author:** Guang Wu, State Key Laboratory of Non-food Biomass Enzyme Technology, National Engineering Research Center for Non-food Biorefinery, Guangxi Key Laboratory of Biorefinery, Guangxi Academy of Sciences, 98 Daling Road, Nanning, Guangxi, 530007, China, E-mail: hongguanglishibahao@yahoo.com

Received December 05, 2012; **Published** April 08, 2013

Citation: Yan S, Wu G (2013) 2-Dimensional HP Folding Structures of Coenzyme Pyrroloquinoline Quinone Synthesis Protein A. 2: 663. doi:10.4172/scientificreports.663

necessary and important to study all the folding structures of PQQA in 2D HP model, which is the aim of this study.

Materials and Methods

Data

A total of 14 PQQAs were obtained from UniProt [14], among them there were 5 same amino acid sequences and two variants of P27532 PQQA [15] so that there were 11 PQQA sequences analyzed in this study. The normalized amino acid hydrophobicity index [16] was obtained from Sigma Aldrich website as shown in table 1.

HP model

HP model classifies amino acids according to either hydrophobic or polar, but does not indicate how to deal with neutral amino acids. Thus the normalized amino acid hydrophobicity index (Table 1) was used to deal with neutral amino acids in PQQA. In fact, this is still not good enough because this normalized amino acid hydrophobicity index is based on the fact that glycine as zero, thus we have to choose glycine either as hydrophobic or as polar. Yet, the amino acid hydrophobicity is pH dependent (Table 1), which leads us to consider the assignment of amino acids of PQQA at two different pH levels. Taken two considerations together, each PQQA has four HP sequences to be operated in HP model. Therefore, totally 44 HP sequences need to operate in HP model as shown in table 2, and each one would theoretically have 94,143,178,827 folding structures.

Results and Discussion

Although the HP model is simple, many problems have yet to solve without mentioning of huge amount of computations, for example, how to deal with neutral amino acids. Certainly, the normalized amino acid hydrophobicity index (Table 1) provides a way to solve this problem but it leads the amount of computations to increase four-fold (Table 2).

In table 2, each PQQA sequence has to convert to HP sequences

At pH 2		At pH 7	
Very Hydrophobic			
L	100	F	100
I	100	I	99
F	92	W	97
W	84	L	97
V	79	V	76
M	74	M	74
Hydrophobic			
C	52	Y	63
Y	49	C	49
A	47	A	41
Neutral			
T	13	T	13
E	8	H	8
G	0	G	0
S	–7	S	–5
Q	–18	Q	–10
D	–18		
Hydrophilic			
R	–26	R	–14
K	–37	K	–23
N	–41	N	–28
H	–42	E	–31

Table 1: Normalized amino acid hydrophobicity index.

PQQA	Classification	Sequence
B3R8D5	Amino acid	MTWTTPAYTELRLGFEITMYIANR
	G=H at pH 2	HHHHHPHHHHHPHHHHHHHHHHPP
	G=P at pH 2	HHHHHPHHHHHPHPHHHHHHHHPP
	G=H at pH 7	HHHHHPHHHPHPHHHPHHHHHHPP
	G=P at pH 7	HHHHHPHHHPHPHPHPHHHHHHPP
B4EHL5	Amino acid	MQWTTPSYTDLRFGFEITMYIANR
	G=H at pH 2	HPHHHPPHHPHPHHHHHHHHHHPP
	G=P at pH 2	HPHHHPPHHPHPHPHHHHHHHHPP
	G=H at pH 7	HPHHHPPHHPHPHHHPHHHHHHPP
	G=P at pH 7	HPHHHPPHHPHPHPHPHHHHHHPP
C5CPT7	Amino acid	MKWETPTATDLRFGFEITMYVSAR
	G=H at pH 2	HPHHHPHHHPHPHHHHHHHHHPHP
	G=P at pH 2	HPHHHPHHHPHPHPHHHHHHHHPP
	G=H at pH 7	HPHPHPHHHPHPHHHPHHHHHPHP
	G=P at pH 7	HPHPHPHHHPHPHPHPHHHHHPHP
P27532 (B0VQD3, B0V498)	Amino acid	MQWTKPAFTDLRIGFEVTMYFEAR
	G=H at pH 2	HPHHPPHHHPHPHHHHHHHHHHHP
	G=P at pH 2	HPHHPPHHHPHPHPHHHHHHHHHP
	G=H at pH 7	HPHHPPHHHPHPHHHPHHHHHPHP
	G=P at pH 7	HPHHPPHHHPHPHPHPHHHHHPHP
P27532 E16D	Amino acid	MQWTKPAFTDLRIGFDVTMYFEAR
	G=H at pH 2	HPHHPPHHHPHPHHHPHHHHHHHP
	G=P at pH 2	HPHHPPHHHPHPHPHPHHHHHHHP
	G=H at pH 7	HPHHPPHHHPHPHHHPHHHHHPHP
	G=P at pH 7	HPHHPPHHHPHPHPHPHHHHHPHP
P27532 Y20F	Amino acid	MQWTKPAFTDLRIGFEVTMFFEAR
	G=H at pH 2	HPHHPPHHHPHPHHHHHHHHHHHP
	G=P at pH 2	HPHHPPHHHPHPHPHHHHHHHHHP
	G=H at pH 7	HPHHPPHHHPHPHHHPHHHHHPHP
	G=P at pH 7	HPHHPPHHHPHPHPHPHHHHHPHP
P55171 (C3K347)	Amino acid	MTWSKPAYTDLRIGFEVTMYFASR
	G=H at pH2	HHHPPPHHHPHPHHHHHHHHHHPP
	G=P at pH2	HHHPPPHHHPHPHPHHHHHHHHPP
	G=H at pH 7	HHHPPPHHHPHPHHHPHHHHHHPP
	G=P at pH 7	HHHPPPHHHPHPHPHPHHHHHHPP
Q3K5R0	Amino acid	MAWTKPAYTDLRIGFEVTMYFASR
	G=H at pH 2	HHHHPPHHHPHPHHHHHHHHHHPP
	G=P at pH 2	HHHHPPHHHPHPHPHHHHHHHHPP
	G=H at pH 7	HHHHPPHHHPHPHHHPHHHHHHPP
	G=P at pH 7	HHHHPPHHHPHPHPHPHHHHHHPP
Q50436	Amino acid	MMWTKPEVTEMRFGFEVTMYVCNR
	G=H at pH2	HHHHPPHHHHHPHHHHHHHHHHPP
	G=P at pH2	HHHHPPHHHHHPHPHHHHHHHHPP
	G=H at pH 7	HHHHPPPHHPHPHHHPHHHHHHPP
	G=P at pH 7	HHHHPPPHHPHPHPHPHHHHHHPP
Q608P4	Amino acid	MRWEKPSYNDMRFGFEVTMYIYNR
	G=H at pH 2	HPHHPPPPHPPHPHHHHHHHHHPP
	G=P at pH 2	HPHHPPPPHPPHPHPHHHHHHHPP
	G=H at pH 7	HPHPPPPPHPPHPHHHPHHHHHPP
	G=P at pH 7	HPHPPPPPHPPHPHPHPHHHHHPP
Q88A80(Q4ZMC5, Q48CT7)	Amino acid	MSWTKPAYTDLRIGFEVTMYFASR
	G=H at pH 2	HPHHPPHHHPHPHHHHHHHHHHPP
	G=P at pH 2	HPHHPPHHHPHPHPHHHHHHHHPP
	G=H at pH 7	HPHHPPHHHPHPHHHPHHHHHHPP
	G=P at pH 7	HPHHPPHHHPHPHPHPHHHHHHPP

G=H at pH 2, Glycine as hydrophobic amino acid at pH 2; G=P at pH 2, Glycine as polar amino acid at pH 2; G=H at pH 7, Glycine as hydrophobic amino acid at pH 7; G=P at pH 7, Glycine as polar amino acid at pH 7.

Table 2: Nine PQQAs, two variants of P27532 PQQA and their HP sequences.

PQQA	G=H at pH 2	G=P at pH 2	G=H at pH 7	G=P at pH 7
B3R8D5	8192 (–13)	564 (–13)	2220 (–12)	1716 (–11)
B4EHL5	1392 (–12)	72 (–12)	1236 (–11)	804 (–10)
C5CPT7	252 (–12)	8192 (–10)	8192 (–8)	8192 (–6)
P27532	3684 (–12)	4152 (–11)	4884 (–10)	8192 (–8)
P27532 E16D	3252 (–11)	1944 (–10)	4884 (–10)	8192 (–8)
P27532 Y20F	3684 (–12)	4152 (–11)	4884 (–10)	8192 (–8)
P55171	6504 (–11)	276 (–11)	204 (–11)	72 (–10)
Q3K5R0	8192 (–12)	492 (–12)	324 (–12)	348 (–11)
Q50436	1200 (–13)	8192 (–12)	1980 (–11)	1740 (–10)
Q608P4	408 (–11)	1692 (–10)	1164 (–9)	1848 (–8)
Q88A80	1020 (–12)	2580 (–11)	1656 (–11)	1320 (–10)

The first number is the number of native states, and the number in parenthesis is the minimal energy.

Table 3: Number of native states according to the minimal energy determined by HP model.

with respect to two pH levels and whether glycine is considered as hydrophobic or as polar. Comparison of HP sequences shows effects of neutral amino acids. For example, difference between G=H and G=P and between pH 2 and pH 7 is the glycine (G) at position 14 and the glutamic acid (E) at positions 10 and 16 with respect to its conversion to either H or P (rows from 3 to 6). Nevertheless, HP sequence can vary to different degree according to a protein composition.

Of 94,143,178,827 possible folding structures for each HP sequence in table 2, it is important to know how many native states a PQQA has. Table 3 shows the number of native states of folding structures with minimal energy of PQQA according to different HP conversions. As seen, PQQA has many native states, for example, 8192 native states are found at pH 2 with glycine assigned as polar, and each native state has the same amount of minimal energy, –13.

It is suggested that a native state represents a thermodynamically stable state corresponding to the global minimum of Gibbs free energy [17], thus, so many native states found by HP model suggest that an enzyme can use different folding structures in a particular enzymatic reaction because each native state can be viewed equally in terms of their minimal energy. The fact that there is more than one native state also suggests the flexibility of folding mechanisms in PQQA.

Figure 1 illustrates 8 native states of B3R8D5 PQQA at pH 7 with glycine assigned as polar because of limitation of space. Each graph in this figure shows how HP model folds a protein in 2D lattice, starting from position 1 to position 24. Such structure constitutes 13 non-sequential H-H connections, and each is a unit of negative energy as 1H-14H, 1H-16H, 1H-18H, 2H-9H, 2H-11H, 3H-8H, 3H-18H, 4H-7H, 4H-19H, 5H-22H, 11H-14H, 17H-20H, and 19H-22H. An interesting point is that the native state is chirally symmetric between four folding structures on the left-hand side and those on the right-hand side, that is, pathways to construct a native state are chiral because they cannot be superimposed in mirror image. Because of symmetry, an enzyme finds its native state more easily through different pathways, which nevertheless minimize the time spending on searching for the native state.

In table 3, the number of native states with their minimal energy is different at different pH levels and through different conversion of glycine as hydrophobic or polar. Figure 2 displays how these two conditions affect the folding structures. First, two pH levels influence amino acid E at positions 4 and 16, i.e. E is converted to H at pH 2 and P at pH 7. Comparing two graphs in middle with two graphs in bottom, 1H-4H, 4H-17H, 9H-16H and 11H-16H do not exist at pH 7. Second, the conversion of G to H or P is related to G at position 14, so it leads to difference on whether 11H-14H and 14H-23H exist by comparing two graphs in left-hand site with two graphs in right-hand site. Actually, those four structures have the same pathway to fold PQQA, but if E in PQQA is located at very edge of HP folding structures, it would be expected to have a less chance construct H-H connections than its current position. Consequently, we would expect to see more dramatic difference if E is located in internal part of configuration of a peptide or protein.

Again, table 3 demonstrates that an enzyme can have many native states, for example, B3R8D5 PQQA has 564 native states when G=P at pH 2 (cell 3, row 2). An intriguing question is whether native states are numerically distinguishable because a non-sequential H-H connection only gives a unit of minimal energy. However, an H-H connection is composed of amino acids with different hydrophobicity, therefore we should furthermore quantify non-sequential H-H connections with normalized amino acid hydrophobicity index (Table 1).

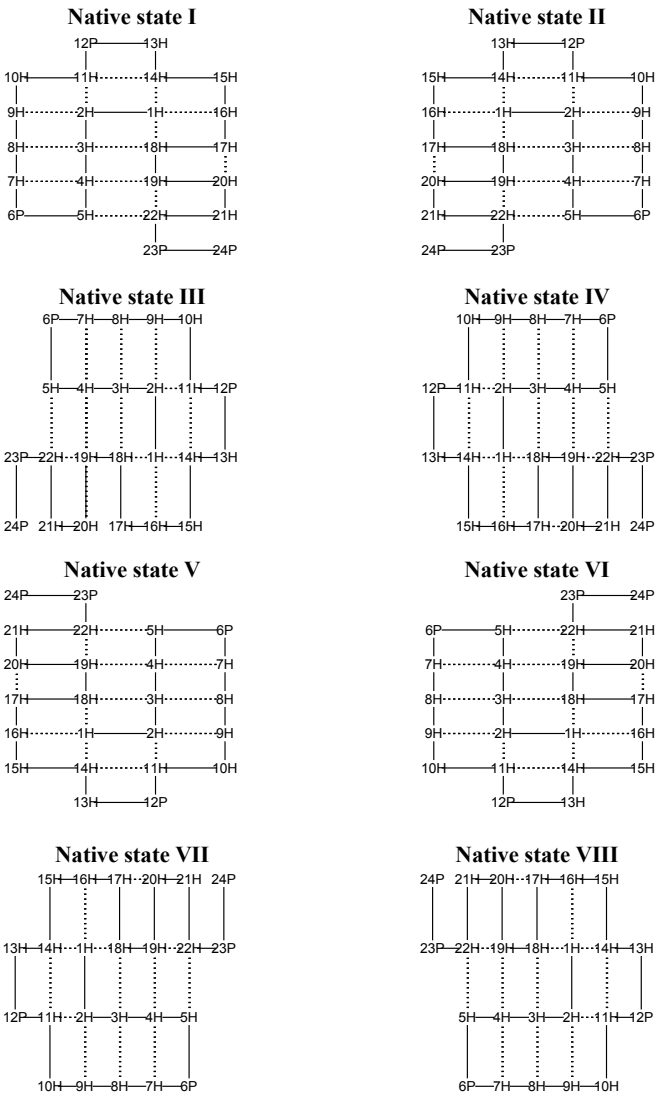

Figure 1: Eight native states of folding structures of B3R8D5 PQQA under HP conversion of G=H at pH 2. The dotted lines are non-sequential H-H connection, which is considered as a unit of negative energy –1, and their sum is the minimal energy –13.

Table 4 lists how to use the normalized amino acid hydrophobicity index to distinguish native states of C5CPT7 PQQA with G=H at pH 2 in terms of sum of hydrophobicity index of amino acids in H-H connections. The detailed computation is as follows: the middle graph in left-hand site in figure 2 shows 12 H-H connections, 1H-4H, 3H-18H, 4H-17H, 5H-8H, 8H-17H, 9H-16H, 11H-14H, 11H-16H, 14H-23H, 15H-18H, 15H-20H, and 20H-23H, which correspond to 1M-4E, 3W-18T, 4E-17I, 5T-8A, 8A-17I, 9T-16E, 11L-14G, 11L-16E, 14G-23A, 15F-18T, 15F-20Y, and 20Y-23A. These amino acid connections can be quantified using table 1, for example, M=74 and E=8 at pH 2 in table 1, so 1H-4H is 1M-4E, and their sum is 74+8=82. In this manner, native states can be distinguishable in table 4.

Sum of hydrophobicity index of H-H connections	Number of Native States
1066	12
1076	12
1084	24
1098	48
1102	12
1104	12
1108	48
1112	12
1134	24
1136	24
1144	24

Table 4: Number of native states and their numerical distinctions determined by the normalized amino acid hydropHobicity index for C5CPT7 PQQA with G=H at pH 2.

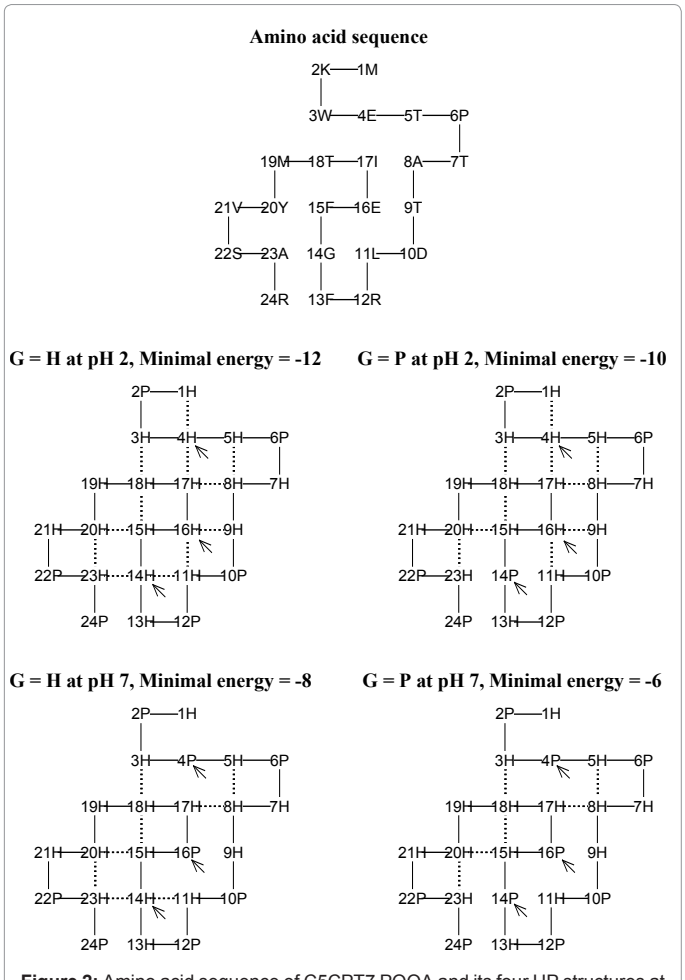

Figure 2: Amino acid sequence of C5CPT7 PQQA and its four HP structures at pH 2 and pH 7 with glycines assigned as hydrophobic as well as polar.

Figure 3 furthermore explains how different native states should have different sum of hydrophobicity index. In this figure, four different native states have the same minimal energy –11 although their structures are different. This is because they have the same number non-sequential H-H connections. As abovementioned, the composition of amino acids for those H-H connections is different, so they should be distinguishable (Figure 3).

In the past, few studies were directed to real-life case when using

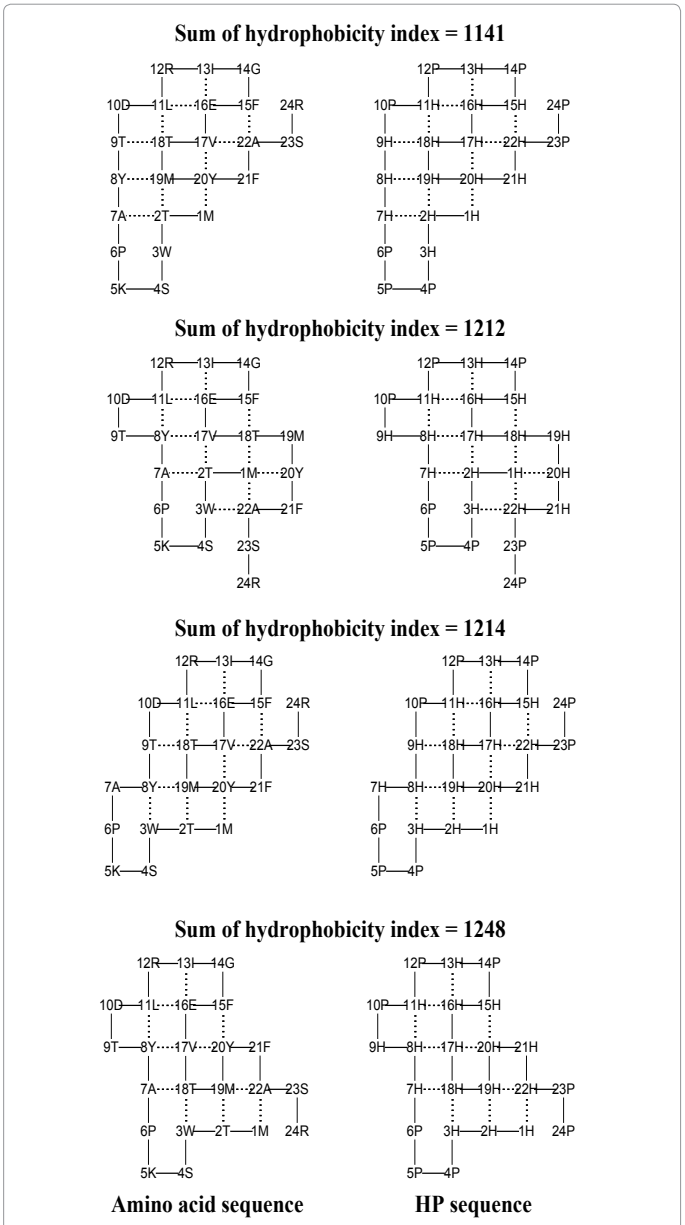

Figure 3: Distinguishing of native states of P55171 PQQA with G=H at pH 2 using the normalized amino acid hydrophobicity index. All native states have a minimal energy –11 by summing dotted lines but different sums of hydrophobicity index of H-H connections. Left-hand site: amino acid sequence; Right-hand site: HP sequence.

HP model [18-20], however, this does not suggest that HP model is useless. The very nature of difficulty in application of HP model is the huge amount of computations. In this study, an important implication is to increase the number of H in order to decrease the minimal energy in a native state, which may lead better kinetic effect for enzyme in enzymatic reactions. Practically, a hydrophobic amino acid can be used to substitute a neutral or hydrophilic amino acid to get a native state with lower minimal energy if this native state concerns enzymatic reactions.

In conclusion, the results from this study suggest possible ways to modify PQQA that we can either modify PQQA via replacing polar amino acids with hydrophobic amino acids or modify PQQA via replacing amino acids according to the normalized amino acid hydrophobicity index.

Acknowledgements

This study was partly supported by Guangxi Science Foundation (2010GXNSFF013003, 10-046-06 and 11-031-11) and by Guangxi Academy of Sciences (11YJ24KY01).

References

1. http://www.ebi.ac.uk/chebi/advancedSearchFT.do?searchString=58442.
2. Gomelsky M, Biville F, Gasser F, Tsygankov YD (1996) Identification and characterization of the *pqqDGC* gene cluster involved in pyrroloquinoline quinone production in an obligate methylotroph *Methylobacillus flagellatum*. FEMS Microbiol Lett 141: 169-176.
3. Dill KA (1985) Theory for the folding and stability of globular proteins. Biochemistry 24: 1501-1509.
4. Lau KF, Dill KA (1989) A lattice statistical mechanics model of the conformational and sequence spaces of proteins. Macromolecules 22: 3986-3997.
5. Berger B, Leighton T (1998) Protein folding in the hydrophobic-hydrophilic (HP) model is NP-complete. J Comput Biol 5: 27-40.
6. Unger R, Moult J (1993) Finding the lowest free energy conformation of a protein is an NP-hard problem: proof and implications. Bull Math Biol 55: 1183-1198.
7. http://en.wikipedia.org/wiki/Millennium_Prize_Problems.
8. Unger R, Moult J (1993) Genetic algorithms for protein folding simulations. J Mol Biol 231: 75-81.
9. Ramakrishnan R, Ramachandran B, Pekny JF (1997) A dynamic Monte Carlo algorithm for exploration of dense conformational spaces in heteropolymers. J Chem Phys 106: 2418-2826.
10. Shmygelska A, Hoos HH (2005)An ant colony optimisation algorithm for the 2D and 3D hydrophobic polar protein folding problem. BMC Bioinformatics 6: 30.
11. Bastolla U, Frauenkron H, Gerstner E, Grassberger P, Nadler W (1998) Testing a new Monte Carlo algorithm for protein folding. Proteins 32: 52-66.
12. Beutler TC, Dill KA (1996) A fast conformational search strategy for finding low energy structures of model proteins. Protein Sci 5: 2037-2043.
13. Lyngsø RB, Pedersen CN (2000) RNA pseudoknot prediction in energy-based models. J Comput Biol 7: 409-427.
14. UniProt Consortium (2010) The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res 38: D142-D148.
15. Goosen N, Huinen RG, van de Putte P (1992) A 24-amino-acid polypeptide is essential for the biosynthesis of the coenzyme pyrrolo-quinoline-quinone. J Bacteriol 174: 1426-1427.
16. Sigma-Aldrich (2011) Amino Acids Reference Chart.
17. Govindarajan S, Goldstein RA (1998) On the thermodynamic hypothesis of protein folding. Proc Natl Acad Sci USA 95: 5545-5549.
18. Yan S, Wu G (2012) Analysis on folding of misgurin using two-dimensional HP model. Proteins 80: 764-773.
19. Yan S, Wu G (2012) Detailed folding structures of M-lycotoxin-Hc1a and its mutageneses using 2-dimensional HP model. Mol Simul 38: 809-822.
20. Yan S, Wu G (2012) Detailed folding structures of Kappa-conotoxin RIIIJ and its mutageneses obtained from 2-dimensional HP model. Protein Pept Lett 19: 567-572.