ELSEVIER

Journal of Molecular Catalysis B: Enzymatic 6 (1999) 133–143

JOURNAL OF MOLECULAR CATALYSIS B: ENZYMATIC

Review

Amino acid substitutions affecting protein solubility: high level expression of *Streptomyces clavuligerus* isopenicillin N synthase in *Escherichia coli* [1]

Janet Sim, Tiow-Suan Sim *

Department of Microbiology, Faculty of Medicine, National University of Singapore, 10 Kent Ridge Cresent, Singapore 119260, Singapore

Received 27 January 1998; revised 24 February 1998; accepted 25 February 1998

Abstract

Modification of specific cultivation conditions, the choice of promoters, host strains and temperatures used for expression have often been exploited to optimize protein folding for soluble production. However, such overexpression of foreign proteins, especially in *Escherichia coli*, often results in inclusion body formation. Besides, when a protein's primary sequence is altered by substitutions at certain amino acid sites, the expressed protein may be rendered insoluble. At present, the mechanism by which such replacements affect solubility is not entirely clear. In this review, it is observed that protein insolubility is not totally dependent on parameters such as hydrophobicity, charge and identity of the amino acid substitutions. Neither is it plainly related to the biophysical properties of the mutated proteins, such as hydropathicity scores and pI values. However, a survey of reported data on ten proteins suggests that increasing the hydrophilicity of solvent-exposed residues could increase solubility and vice versa. In addition, results obtained from computational analysis and expression studies of isopenicillin N synthase (IPNS) mutants indicate an apparent causal relationship between secondary structure predictions and expression of soluble proteins. Hence, specific amino acid substitutions affecting secondary structure predictions and thereby protein folding, are expected to have a greater influence on protein solubility than a trivial assessment of other biophysical parameters.

Keywords: Amino acid substitutions; Solubility; Solvent accessibility profiles; Secondary Structure; Folding

1. Introduction

The need to provide new and more potent antibiotics to combat emerging resistant or 'new' pathogens depends, to a large extent, on the genetic engineering of enzymes involved in the β-lactams biosynthesis. Such pursuits include expanding the substrate range and improving the biological activity of key enzymes with the implicit requirement of achieving high level expression of soluble enzymes.

To this end, the enzyme in the β-lactam biosynthetic pathway that has been extensively studied is isopenicillin N synthase (IPNS). The

* Corresponding author.

[1] Dedicated to Professor Hideaki Yamada in honor of his 70th birthday.

1381-1177/99/$ - see front matter
PII: S1381-1177(98)00072-1

intense focus on IPNS stems from its importance as the key enzyme found in all cephalosporin and penicillin producers that catalyze the oxidative cyclization of the tripeptide δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV) [1]. With its unusually relaxed substrate specificity, IPNS demonstrates great potential to be exploited via protein engineering to produce new potent β-lactam antibiotics. So far, IPNS genes have been cloned from various fungal and bacterial sources [2] to investigate the possibility of producing a large amount of protein necessary for enzyme characterization via expression in *Escherichia coli*.

It is commonly reported that high level expression of foreign proteins in *E. coli* results in the accumulation of the product as insoluble aggregates in the cytoplasm or periplasm of the producing bacteria in the form of 'inclusion bodies' (IBs) [3]. Although many proteins have been purified from IBs and some have yielded crystal structures [4], IB formation is often considered undesirable, as there are still some intractable problems associated with IB protein purification schemes (reviewed in Refs. [3,5,6]). Thus, it would be important to develop strategies for efficient production of *soluble* proteins in *E. coli* for enzymatic study.

Thus far, there have been numerous reports on the expression of IPNS isozymes in the heterologous host *E. coli*. From the analysis of the gene expression studies [7,8], it is noted that fungal IPNS is generally reported to be more soluble than the bacterial IPNS. Nevertheless, overproduction of recombinant IPNS isozymes made possible via manipulation of the expression systems and cultivation conditions for transformed *E. coli* has facilitated purification of IPNS proteins to near homogeneity using relatively simple procedures [8–10]. These have provided the means for biochemical and biophysical analyses of IPNS. In the numerous site-directed mutagenesis experiments designed specifically to locate the active site residues [10–14], numerous IPNS mutants from both fungal and bacterial sources have been generated. However, there was no emphasis on whether the site changes introduced into the wild type IPNS sequence has affected the protein folding and solubility of the mutants in the various reports.

Up to the present, mutations that affect aggregate formation or soluble protein production have been observed in many systems [15,16]. However, the genetic code of how a specific amino acid sequence is first directed to fold into protein intermediates and finally into the well-defined three-dimensional native conformation in split seconds remains an unfathomable mystery.

Therefore, this paper aims to investigate the amino acid sites that contribute to the efficiency of the folding process, and thereby the solubility in different proteins. Included in this analysis is a study on mutants of *Streptomyces clavuligerus* IPNS (scIPNS) obtained from site-directed and random mutagenesis. There appears to be a close relationship between the predicted secondary structures of proteins and the tendency to form soluble proteins.

2. Amino acid substitutions influencing protein solubility

Protein solubility has been defined by the ability of soluble, polar residues to interact with water such that the rest of the protein could fold properly to give a well-defined active conformation [17]. Such stable interactions could most probably have been disrupted in several mutant proteins reported, where small changes in the primary sequence have caused dramatic changes in wild type solubility. Despite numerous attempts to correlate mutations with changes in solubility [15–17], the mechanism by which how amino acid substitutions alter solubility is still poorly understood.

Comparative studies of altered sequences in other mutated proteins showed that identical replacements at different amino acid positions

do not always affect inclusion body formation to the same extent, suggesting that insolubility does not invariably depend on parameters such as hydrophobicity or charge, but may be more importantly related to protein structure [18,19]. Thus, to analyze the relationship between the nature of amino acid substitutions, the spatial location of the residues altered and the solubility pattern, point mutations in 10 recombinant proteins that have altered solubility are divided into two groups (Table 1) based on the structural location of the replacements.

2.1. Substitution of solvent-exposed residues

In an extensive analysis of the evolution of globin sequences from many species, Perutz and co-workers found that there exists strong dispo-

Table 1
Amino acid substitutions influencing protein solubility

Substitutions at surface exposed residues

Name of enzyme	Amino acid residues mutated	Hydrophobicity[a]	Solubility[b]	References
S1 dihydrofolate reductase from *Staphylococcus aureus*	Asn48Glu	No change	↑	[18]
	Asn130Asp	No change	↑	
	Asn48Glu/Asn130Asp	No change	↑	
Catalytic core domain of HIV type 1 integrase	Val65Lys	↑	↑	[20]
	Phe185Lys	↑	↑	
Human α1-proteinase inhibitor	Met358Arg/Met351GLu	↑/↑	↑	[21]
	Met358Arg/Thr345Leu	↑/↓	↓	
	Met358Leu	↓	↓	
Human interleukin-1β	Lys97Arg	↑	↑	[22]
	Lys97Gly	↓	↓	
	Lys97Val	↓	↓	
Human thymidylate synthase	Cys13Glu	↑	↑	[23]
	Pro139Asp	↑	↑	
	Leu140Lys	↑	↑	
Human medium-chain acyl-CoA dehydrogenase[c]	Lys329Glu	No change	↓	[19]
E. coli UMP-kinase[d]	Asn159Asp	No change	↓	[24]
E. coli maltose-binding protein[e]	Gly32Asp/Ile33Pro	↑/↑	↓	[25]

Substitutions at core of protein

Name of enzyme	Amino acid residues mutated	Hydrophobicity[a]	Solubility	Stability	References
Colicin A	Trp140Lys	↑	↓	↓	[26]
	Trp140Leu	↓	↓	↓	
	Trp140Cys	↓	↓	↓	
	Trp140Lys/Lys113Phe	↑/↓	↑	↑	
	Trp140Leu/Lys113Phe	↓/↓	↑	↑	
	Trp140Cys/Lys113Phe	↓/↓	↑	↑	
Human interleukin 1-β	Leu10Asn	↑	↓	↓	[22]
	Leu10Asp	↑	↓	↓	
	Leu10Thr	↑	↓	↓	

[a] Hydrophobic indices are obtained from Kyte and Doolittle [27]. Substitutions that increase, decrease or do not alter the hydro-philicity of the sites are denoted by '↑','↓', and 'no change' respectively.
[b] Increased or reduced solubility of the respective mutants constructed compared to wild type are designated by symbols ↑ and ↓, respectively.
[c] Lys329 of human MCAD is involved in making intersubunit contacts, thus mutation at this site may lead to aggregation.
[d] Asn159 of *E. coli* UMP-kinase is involved in the formation of salt-bridges; disruption of this salt bridge would cause conformational changes that might result in insolubility.
[e] The authors proposed that a proline residue at position 33 (located in a turn of maltose-binding protein) could introduce a conformation strain in the folded state.

sition against large patches of hydrophobic residues on the surface of globin sequences [28]. This may reflect evolutionary constraints imposed by protein solubility as surface-exposed hydrophobic residues may interact unfavorably with aqueous solvents, presumably leading to aggregation. On the other hand, it has been suggested that replacement of surface residues to increase hydrophilicity could increase the solubility [17]. However, collective analysis of experimental examples to affirm this proposition was not done. Therefore, it would be useful to investigate how amino acid substitutions at sites experimentally determined to be located at the surface of eight proteins can influence the solubility of the mutated proteins.

In accordance to what is observed in nature for globin proteins, increasing the distribution of surface charges of four of the proteins analyzed, i.e., replacing solvent exposed hydrophobic residues in contact with the solvent to more hydrophilic ones, has successfully improved the solubility of the specifically constructed mutants and vice versa. The four proteins are, namely, the catalytic core domain of HIV type 1 integrase [20], human α1-proteinase inhibitor [21], human interleukin-1β [22] and human thimidylate synthase [23]. For instance, separate substitution of hydrophobic residues, Val65 and Phe185, of catalytic core domain of HIV integrase by the hydrophilic lysine amino acid has improved the solubility of the wild type protein. The reverse is also true, as replacement of hydrophilic lysine residue at position 97 on the surface of human interleukin-1β by hydrophobic residues, glycine and valine, has decreased the level of soluble protein made. In the case of S1 dihydrofolate reductase (DHFR) [18], exploitation of surface residue replacement to improve solubility was carried out by substitution of asparagine residue at positions 48 and 130 by the negatively charged glutamine and aspartate amino acids, respectively. The authors have reported success in enhancing soluble protein production by increasing the negative charge distribution on the surface of S1 DHFR.

However, this notion was not applicable to three of the examples analyzed, viz., medium-chain acyl-CoA dehydrogenase (MCAD) [19], UMP-kinase [24] and maltose-binding protein [25], where mutations were engineered at surface residues proposed to be responsible for structure formation or stability (refer to legend of Table 1). For example, when the positively charged lysine residue located at position 329 of MCAD was mutated to the negatively charged glutamine with the same hydrophobic index values, it is assumed that the solubility of the mutant protein would either be improved or remain the same as the wild type level. Unexpectedly, the amino acid substitution has resulted in a drastic reduction in the level of soluble protein made probably because this site change may interfere with the assembly of the native tetrameric form, thus leading to aggregation.

In conclusion, it seems that a practical approach to improve the solubility of the recombinant protein expressed could be achieved by selective engineering of solvent-exposed residues to less hydrophobic or negatively charged ones. However, replacement of surface residues with an intrinsic function, e.g., those involved in maintaining the stability of loops (especially proline in turns), hydrophobic packing, formation of pertinent hydrogen bonding and salt bridges, should be avoided.

2.2. Substitution of core residues

With rare exceptions, the core of proteins must be efficiently packed and remain hydrophobic [29–31]. Cavities are rare in the hydrophobic center of most natural proteins, and only small volume changes are tolerated, as certain packing density are mandatory to maintain the stability of the native structure. Besides, substitutions of the core residues with polar and charged residues usually have deleterious effect on the protein structure unless complementary packing could still be maintained without assuming a strained conformation.

Although the above observations relate more to the stability of protein, however, we would like to examine whether destabilization mutations of the hydrophobic core would also simultaneously affect the solubility of the mutant protein. In the analysis of the human interleukin-1β mutants [22], it was reported that when three polar residues, asparagine, aspartate and threonine, were separately introduced into the buried site at Leu10, the respective single mutants constructed have reduced stability and are also expressed in lower levels in the soluble fraction as compared to the wild type protein. In another study [26], unfavorable substitution at position 140 of colicin A has created a hydrophobic cavity that has concomitantly resulted in a reduction in the level of the soluble protein produced. A single mutation, substitution of lysine 113 with phenyalanine, can fill up the cavity created previously, thereby restoring the solubility of the various mutants.

Although analysis of the mutagenesis studies of the above two proteins revealed that amino acid substitutions that interfere with the proper packing of the core of the native protein, i.e., reduction in stability, would invariably decrease the solubility of the resultant protein, more examples are needed to validate that the observation is a norm rather than an exception.

3. Amino acid substitutions that alter *S. clavuligerus* IPNS solubility

Previous studies have reported that although bacterial scIPNS (*S. clavuligerus* IPNS) was expressed predominantly in the insoluble form at 37°C, soluble form (~29% of total soluble protein) could be obtained when the cultivation temperature was lowered to 25°C [8]. However, we have isolated mutants of scIPNS obtained from site-directed and random mutagenesis that showed reduced production of soluble protein even at 25°C (unpublished). Sequencing analysis was carried out to ascertain that the site changes have occurred in the primary structure. This was followed up by computational analysis to predict the structural locations of the mutations in the IPNS variants.

3.1. Expression studies and sequence analysis

In the functional analysis of two site-directed mutants (Asp214Ala and Gln328Leu) (Table 2), it was observed that single site change introduced at amino acid residue 214 of scIPNS gene resulted in a reduced level of expression of the mutant protein in the soluble fraction, whereas the Gln328Leu mutant appears to have the same solubility as the wild type protein [8].

In addition, we have also isolated seven scIPNS clones that showed variable solubility when expressed under the same condition (Table 2). These clones were generated in an attempt to subclone scIPNS genes, amplified using low fidelity *Taq* polymerase, for high level expression in *E. coli* [8]. As it is well-documented that *Taq* polymerase exhibits a high nucleotide misincorporation rate [34–36], the full IPNS sequences of each of the seven clones were determined to identify whether any misincorporations have been introduced into the wild type gene sequence during polymerase chain reaction (PCR) which could account for the altered phenotypes. Sequencing results showed that the mutants were found to harbor, in addition to the intended change at the fourth nucleotide position, nucleotide misincorporations at different positions throughout the entire gene.

The nucleotide positions where scIPNS gene has been mutated, the corresponding amino acid residue changes and the expression levels in the soluble fraction of the various clones are listed in Table 2. It appears that these clones with altered solubility all possess distinct mutations. In clones 36, 43, 48 and D214A, mutations at no more than three positions of scIPNS protein (~329 amino acid residues) have caused the mutant enzyme to become predominantly aggregated in the inclusion bodies. This observation unveils the fragile relationship between protein sequence and structure, such that a slight change

Table 2
Amino acid substitutions that can affect the solubility of scIPNS

Mutant	Nucleotide substituted	Corresponding amino acid residue changed[a]	Solvent accessibility profile of mutated residue predictedy by PHDacc[b]	Hydrophilicity[c]	Predicted secondary structure element of mutated site[d]	Number of sites where secondary structure elements are altered[e]		GRAVY score of mutant[f]	pI of mutant[f]	Expression level (expressed as % of total soluble protein)[g]
						SOPM	nnPred			
52	C4G	Pro2Ala	×	↓	coil	3	0	−3.249	5.18	22
34	A932G	Tyr311Cys	×	↓	α-helix	12	2	−3.134	5.18	26
36	C389T	Pro130Leu	×	↓	coil	22	11	−3.116	5.25	5
	A680G	Gln227Arg	×	↑	α-helix					
42	A596G	Lys199Arg	×	↑	coil	6	0	−3.365	5.25	23
43	C194G	Ala65Gly	Exposed	↑	α-helix	19	3	−3.222	5.24	7
	A536G	Asp179Gly	Exposed	↓	coil					
47	A598G	Thr200Ala	×	↓	coil	17	0	−3.216	5.18	17
	G871T	Val291Phe	×	↑	β-sheets					
48	A689T	Gln230Leu	×	↓	β-sheets	29	12	−3.191	5.18	6
	T965C	Leu322Pro	Buried	↑	β-sheets					
D214A	A641G	Asp214Ala	Exposed	↓	coil	18	7	−3.191	5.24	3
Q328L	A983T	Gln328Leu	Exposed	↓	coil	1	0	−3.131	5.18	23

[a] By purportedly changing the fourth nucleotide by PCR, the second amino acid residue would be changed from proline to alanine. All random mutants (34, 36, 42, 43, 47, 48) contain Pro2Ala in addition to the mutations listed in the table.
[b] × means that no prediction is made for the solvent accessibility for that particular mutated residue as the reliability index of prediction is too low.
[c] Hydrophobicity indices are obtained from Kyte and Doolittle [27]. Substitutions that increase or decrease the hydrophilicity of the mutated site are denoted by ↑ and ↓ respectively.
[d] The secondary structure elements stated for all the mutated residues are predicted from SOPM [32].
[e] The numbers in the table were obtained by enumerating the number of secondary structure elements in the respective mutants that are different from that of the wild type secondary structure predicted by SOPM [32] and nnPred [33].
[f] GRAVY and pI of the various mutants are predicted from ProtParam tool. (website: http://expasy.hcuge.ch/cgi-bin/)
[g] The percentage of various scIPNS mutants expressed in the soluble fraction was obtained via densitometric scanning of the SDS-PAGE analysis (gel not shown).

in the primary structure could topple the scale between folding and unfolding, resulting in aggregate formation. Thus, it would be interesting to determine if there is any rule governing the insolubility of scIPNS, such that a specific amino acid sequence composition could be translated to a measurable property of the protein which would in turn define the solubility profile.

3.2. Characterization of IPNS variants via computational analysis

To date, only the crystal structure of IPNS from the fungal species, *Aspergillus nidulans* (aIPNS), has been elucidated [37,38]. However, the coordinates of the resolved structure may not be readily accessible from Protein Data Bank (PDB), an archive of experimentally determined three-dimensional structure of biological structures. Therefore, the lack of such structural information precludes the use of homology modeling methods to build a model for the closely related bacterial *S. clavuligerus* IPNS. Thus, to gain insight into how specific site alterations in the various scIPNS mutants have caused the protein to aggregate, protein analysis programs are used initially to study whether the mutated residues are buried or exposed in scIPNS and the effect of mutations on the secondary structures of the enzyme.

3.2.1. Solvent accessibility profiles

A computer program based on neural network system (PHDacc) [39] was used to predict the relative solvent accessibility profile of the mutated residues in scIPNS variants. These predictions were then used to study whether the generally observed reduced solubility in the mutants was due to unfavorable substitutions at the surface or core of IPNS in compliance with the 'general rules' defined earlier after examination of the mutations in other recombinant proteins.

The changes in solvent accessibility and hydrophilicity profiles of 13 mutations are listed in Table 2. Complete analysis was impeded by the relatively small number (5 out of 13) of mutated sites that was predicted with high confidence by the program. Nevertheless, analysis results reveal that the 'general rule' proposed in Section 2 could not explain the changes in the solubility profile of scIPNS mutants. For example, substitution of *exposed hydrophilic* aspartate and glutamine residues at positions 214 and 328 respectively with a more *hydrophobic* residue resulted in increased aggregation in D214A mutant, but the solubility of Q328L mutant was unaffected. It is possible that a larger set of data points is needed to make the results obtained from this kind of analysis statistically more significant.

3.2.2. Secondary structure analysis

Two protein secondary structure prediction programs, self-optimized prediction method (SOPM) [32] and neural network prediction program (nnPred) [33], were used in the preliminary analysis of secondary structures of the scIPNS mutants. To validate that the high accuracy claimed by the two methods to predict secondary structure would also apply to the predictions of IPNS structure, a control analysis was done. The secondary structure of aIPNS was determined separately using SOPM and nnPred and subsequently aligned with the secondary structure determined experimentally from the crystal structure of aIPNS. Comparative analysis showed that the two programs could indeed correctly predict at least 65% of the secondary structure elements.

The secondary structure elements of the mutated sites have been listed in Table 2. However, due to the lack of information on the mutated residue(s) responsible for the decreased solubility in some scIPNS triple mutants (36, 43, 47 and 48), it is not possible at present to determine whether specific mutations that result in insolubility are associated with a particular secondary structure element.

SOPM and nnPred generated structures showed that for mutants that had a secondary structure profile different from that of the wild type, the changes in the profiles are not localized in a specific region of the protein. There-

fore, the number of residue sites where secondary structure elements were changed with respect to wild type was enumerated, and the values obtained plotted against solubility data for all the mutants. From Fig. 1, there appears to be an inverse relationship between predicted secondary structure perturbations and solubility. D214A mutant showed a very interesting profile. A single site change at position 214 has caused the mutant protein to have large changes in the secondary structure at many positions (predicted by both SOPM and nnPred) and the data obtained fit into the graph at a position where the protein would be insoluble. This predicted solubility agrees with the experimented expression profile of D214A.

3.2.3. GRAVY and pI scores

The GRAVY (grand average of hydropathicity) scores and pI values obtained for each mutant showed no obvious correlation with the expression in the soluble fraction. In the study of human thymidylate synthase [23], it was reported that the increased solubility observed in some of the mutants was not related to the pI of these enzymes. However, another group noted that the pI values are related to the solubility of dihydrofolate reductase [18]. The reason for the varied relationship between isoelectric points and solubility in different proteins is unknown.

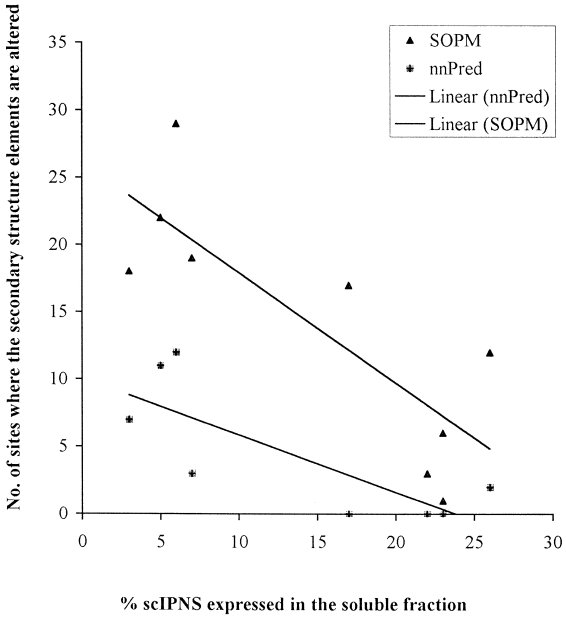

Fig. 1. The graph of percentage pf scIPNS expressed in the soluble fraction vs. number of changes in the secondary structure elements predicted by SOPM and nnPred.

4. Discussion

High level expression of both prokaryotic as well as eukaryotic proteins in *E. coli* often leads to the formation of insoluble aggregates containing the denatured proteins in inclusion bodies (IB) [3]. Moreover, many soluble proteins become insoluble after their primary sequences are altered specifically by site-directed mutagenesis. Although comparative studies [9,11] have shown that the enzyme activities of solubilized IPNS are similar to the soluble form and either one can be used for characterization, it has been reported that solubilization procedures could affect the integrity of refolded IPNS structure, thereby interfering with the interpretation of enzymatic studies. For instance, the circular dichroism (CD) spectra obtained for a solubilized triple mutant constructed during the functional analysis of conserved cysteines in scIPNS, showed lack of ordered structure, indicating that the conformation of the mutant is different from the wild type [12]. Thus, the authors concluded that the reduced activity seen in this triple mutant may be due to inability of the denatured protein to refold properly to give the native structure following solubilization rather than to specific effects of mutation on the active site. In view of the practical problems associated with the biochemical analysis of resolubilized proteins, it appears that soluble protein, with correctly folded native structure, is a better choice not only for characterization of IPNS specific mutants but also for other proteins.

It was mentioned earlier that although scIPNS is predominantly associated with the inclusion bodies when produced at 37°C [40,41], success-

ful overproduction of soluble scIPNS is possible by lowering the cultivation temperature [8,9]. However, in the analysis of the scIPNS mutants, it is observed that slight changes at specific positions of scIPNS gene (which do not affect the total level of IPNS expressed) have rendered the protein insoluble even when the expression was done at 25°C. This observation indicates that these specific site changes appear to have overridden the temperature effect to produce soluble protein, probably by affecting the normal folding pathway of native IPNS.

Point mutations that decreased the solubility of proteins are not novel and are commonly observed in many systems. Mutations that can improve the solubility are more interesting, especially if they can be used as a benchmark for protein engineers to design recombinant proteins that are soluble. Besides, two other parameters, i.e., stability and activity, are also important key factors to consider in producing robust enzymes with altered specificity. However, the relationship between mutation, solubility, stability and activity of proteins in general has not been thoroughly examined and poorly understood.

Although there are extensive experimental studies probing the determinants of stability (reviewed in Refs. [42–44]), an initial effort to obtain a larger collection of examples in this study to analyze the relationship between mutations and solubility have been futile, as there are relatively less investigations studying the latter two parameters. Many reports usually show the effect of amino acid substitutions on stability without simultaneous investigation of the effect on solubility. Although some studies have shown that mutations that increased inclusion body formation had also decreased the stability of the specific mutants [25,26], a few reports had suggested that there exists no strong correlation between the two parameters in relation to mutations [22,45]. Thus, it is possible that the code through which amino acid sequences direct the solubility of polypeptide chains and that of stability in different proteins is not the same. Nevertheless, it would be useful to set up a database consisting of how mutations in various proteins affect properties such as solubility, stability and activity, providing guidelines for protein engineering.

Using only mutations that are explicitly stated in different systems to affect solubility for our analysis, some general rules of thumb have been derived, which could serve as a guideline and not hardcore principles that would be applicable to all proteins. Selective replacements of surface residues with more hydrophilic or negatively charged residue could improve the solubility of proteins and vice versa. However, core residues should only be mutated when detailed structural information is available for the protein in question, as site changes in the core of protein are usually deleterious unless the stability is preserved.

Initial analysis of amino acid substitutions in IPNS variants by considering their solvent exposure profile predicted by PHDacc program in relation to net hydrophobicity index change and the solubility pattern could make no generalization. It is likely that accurate analysis is not possible because in this prediction, the fraction of correctly predicted residue states is only 58%. Information on the solvent exposure pattern of individual amino acids in any protein is most explicitly extrapolated from the crystallographic structure or extensive nuclear magnetic resonance (NMR) study. However, these two methods are technically demanding and require sophisticated machinery implementation. To our knowledge, there are few easy-to-run experiments to investigate this property. Nevertheless, this approach to study amino acid substitutions using computer predictions can still be useful, especially in cases where the crystal structure is not available, or that a 3-D structure is not easily predicted by homology modeling.

Preliminary analysis of the secondary structures of scIPNS mutants using biocomputing programs yield a putative ‘model’ which dictates an inverse relationship between solubility and the extent of change in the secondary struc-

ture predicted. This is interesting, as in vitro protein folding experiments have indicated that stable secondary structure is formed in an early step that provide a framework for subsequent folding [46]. Thus, it is possible that mutations in primary structure, predicted to cause major changes in the number of secondary structure elements, could have affected the formation of the native secondary structure in vivo, and eventually affecting the distribution of the expressed mutant in the soluble and insoluble forms.

Although both prediction programs, SOPM and nnPred, predict the same structural profiles for the various mutants that proposed the same relationship between predicted structures and solubility, experimental evidence is required to support the putative model. For example, site-directed mutagenesis experiments could be done to create mutants with specific site changes to test whether alterations in secondary structure complement changes in biophysical properties.

All the scIPNS mutants analyzed either produced the same or reduced level of soluble protein compared to the wild type at 25°C, indicating that mutations that can increase the solubility of scIPNS are perhaps rare. Although precise amino acid residue(s) responsible for the decreased solubility in some of the mutants has not been defined, there appears to be discrete regions of the primary sequence that codes for solubility or inclusion body formation. For example, amino acid residues 2, 199, 311 and 328 probably do not play a part in soluble protein formation, as substitutions at these locations have not altered the percentage of scIPNS expressed in the soluble fraction. However, aspartate residue at position 214 seems to be important for maintaining the native structure such that mutation at this site has caused D214A to be aggregated. Therefore, such mutants studied could be used as a framework to design experiments to specifically locate sites important for enzyme solubility.

Perhaps one approach to do so would be to generate a larger library of mutants by a repeated PCR random mutagenesis using *Taq* polymerase. However, this method is labor-intensive and time-consuming, as mutants with more than one substitution need to be characterized further. Another option would be to exploit the evolution of enzymes in nature. For example, in the family of IPNS isozymes, fungal IPNS appears to have evolved to possess higher solubility than bacterial IPNS when expressed in *E. coli*. Since the percentage relatedness between the fungal and bacterial IPNS at the nucleotide level is 62–79% and 57–63% at the amino acid level [2], it is possible that the remaining part of the genes or proteins that are not homologous may hypothetically provide the discerning factors responsible for the solubility of expressed proteins. Such regions could be targeted for mutagenesis to determine experimentally, regions important for the proper folding of IPNS.

In mutagenesis experiment to improve catalytic activity/substrate specificity of enzymes, amino acid residues that are critical for proper folding to produce native, soluble protein with correct three-dimensional structure should not be replaced. Therefore, breaking the code for how primary sequence of proteins defines solubility is very important, as it would be a sine qua don for successful engineering of enzymes.

5. Conclusion

Approaches used in this study to analyze mutations in scIPNS would also be useful for other proteins. Systematic replacement of surface/core residues by biocomputational and experimental techniques may be a useful strategy to improve the solubility of proteins to facilitate further structural and biochemical studies.

Acknowledgements

The authors would like to thank Paxton Loke for assistance in the construction and expression studies of the D214A and Q328L scIPNS mu-

tants. This work was supported by National University of Singapore Research Grant RP950390/N and RP950366.

References

[1] J.E. Baldwin, E.P. Abraham, Natl. Prod. Rep. 5 (1988) 129.
[2] J.F. Martin, S. Gutiérrez, Antonie van Leeuwenhoek 67 (1995) 181.
[3] F.A.O. Marston, Biochem. J. 240 (1986) 1.
[4] K. Nagai, H.C. Thogersen, B.F. Luisi, Biochem. Soc. Trans. 16 (1988) 108.
[5] B. Fischer, I. Sumner, P. Goodenough, Biotechnol. Bioeng. 41 (1993) 3.
[6] G. Georgiou, G. Bowden, in: Prokop, A., Bajpai, R.K., Ho, C.S., (Eds.), Recombinant DNA Technology and Applications, McGraw-Hill, USA, 1991, 333.
[7] T.-S. Sim, S.H.D. Tan, Biochem. Mol. Biol. Int. 35 (1995) 1069.
[8] B.J. Sim, D.S.H. Tan, X. Liu, T.-S. Sim, J. Mol. Cataly. B: Enzyme 2 (1996) 71.
[9] M. Durairaj, S.E. Jensen, J. Ind. Microbiol. 16 (1996) 197.
[10] S.H.D. Tan, T.-S. Sim, J. Biol. Chem. 271 (1996) 889.
[11] I. Borovok, O. Landman, R. Kreisberg-Zakarin, Y. Aharonowitz, G. Cohen, Biochemistry 35 (1996) 1981.
[12] M. Durairaj, B.K. Leskiw, J.E. Susan, Can. J. Microbiol. 42 (1996) 870.
[13] M. Sami, T.J.N. Brown, P.L. Roach, C.J. Schofield, J.E. Baldwin, FEBS lett. 405 (1997) 191.
[14] O. Landman, I. Bovorok, Y. Aharonowitz, G. Cohen, FEBS lett. 405 (1997) 172.
[15] C.H. Schein, Curr. Opin. Biotechnol. 4 (1993) 456.
[16] C.H. Schein, Curr. Opin. Biotechnol. 2 (1991) 746.
[17] C.H. Schein, Biotechnology 8 (1990) 308.
[18] G.E. Dale, C. Broger, H. Langen, A. D'Arcy, D. Stüber, Protein Eng. 7 (1994) 933.
[19] N. Gregersen, B.S. Andresen, P. Bross, V. Winter, N. Rüdiger, S. Engst, E. Christensen, D. Kelly, A.W. Strauss, S. Kolvraa, L. Bolund, S. Ghisla, Hum. Genet. 86 (1991) 545.
[20] T.M. Jenkins, A.B. Hickman, F. Dyda, R. Ghirlando, D.R. Davies, R. Craigie, Proc. Natl. Acad. Sci. U.S.A. 92 (1995) 6057.
[21] A.J. Schulze, E. Degryse, D. Speck, R. Huber, R. Bischoff, J. Biotechnol. 32 (1994) 231.
[22] B.A. Chrunyk, J. Evans, J. Lillquist, P. Young, R. Wetzel, J. Biol. Chem. 268 (1993) 18053.
[23] H.E. McElroy, G.W. Sisson, W.E. Schoettlin, R.M. Aust, J.E. Villafranca, J. Cryst. Growth 122 (1992) 265.
[24] L. Serina, N. Bucurenci, A.-M. Gilles, W.K. Surewicz, H. Fabian, H.H. Mantsch, M. Takahashi, I. Petrescu, G. Batelier, O. Barzu, Biochemistry 35 (1996) 7003.
[25] J.-M. Betton, M. Hofnung, J. Biol. Chem. 271 (1996) 8046.
[26] J. Izard, M.W. Parker, M. Chartier, D. Duche, D. Baty, Protein Eng. 7 (1994) 1495.
[27] J. Kyte, R.F. Doolittle, J. Mol. Biol. 157 (1982) 105.
[28] D. Bashford, C. Chothia, A.M. Lesk, J. Mol. Biol. 196 (1987) 199.
[29] F.E. Cohen, D.P. Hearst, Protein Engineering: Principles and Practice, Wiley-Liss, 1996, 33.
[30] M.H.J. Cordes, A.R. Davidson, R.T. Sauer, Curr. Opin. Struct. Biol. 6 (1996) 3.
[31] J.U. Bowie, J.F.R. Olson, W.A. Lim, R.T. Sauer, Science 247 (1990) 1306.
[32] C. Geourjon, G. Deléage, Protein Eng. 7 (1994) 157–164.
[33] D.G. Kneller, F.E. Cohen, R. Langridge, J. Mol. Biol. 214 (1990) 171.
[34] K.R. Tindall, T.A. Kunkel, Biochemistry 27 (1988) 6008.
[35] K.A. Eckert, T.A. Kunkel, Nucleic Acids Res. 18 (1990) 3739.
[36] P.D. Ennis, J. Zemmour, R.D. Salter, P. Parham, Proc. Natl. Acad. Sci. U.S.A. 87 (1990) 2833.
[37] P.L. Roach, I.J. Clifton, V. Fütöp, K. Harlos, G.J. Barton, J. Hadju, I. Anderson, C.J. Schofield, J.E. Baldwin, Nature 375 (1995) 700.
[38] P.L. Roach, I.J. Clifton, C.M.H. Hensgens, N. Shibata, C.J. Schofield, J. Hadju, J.E. Baldwin, Nature 387 (1997) 827.
[39] B. Rost, C. Sander, Proteins 20 (1994) 216.
[40] M. Durairaj, J.L. Doran, S.E. Jensen, Appl. Environ. Microbiol. 58 (1992) 257.
[41] J.L. Doran, B.K. Leskiw, A.K. Petrich, D.W.S. Westlake, S.E. Jensen, J. Ind. Microbiol. 5 (1990) 197.
[42] D. Shortle, J. Biol. Chem. 264 (1989) 5315.
[43] J.M. Sturtevant, Curr. Opin. Struct. Biol. 4 (1994) 69.
[44] E. Querol, J.A. Perez-Pons, A. Mozo-Villarias, Protein Eng. 9 (1996) 265.
[45] E. Bianchi, S. Venturini, A. Pessi, A. Tranmontano, M. Sollazzo, J. Mol. Biol. 236 (1994) 649.
[46] M.-J. Gething, J. Sambrook, Nature 355 (1992) 33.