the FEBS Journal

Using directed evolution to improve the solubility of the C-terminal domain of *Escherichia coli* aminopeptidase P

Implications for metal binding and protein stability

Jian-Wei Liu[1], Kieran S. Hadler[2], Gerhard Schenk[2] and David Ollis[1]

1 Research School of Chemistry, Australian National University, Canberra, Australia
2 School of Molecular and Microbial Sciences, University of Queensland, Brisbane, Australia

Keywords
directed evolution; domain; fusion; metalloprotein; protein solubility

Correspondence
J.-W. Liu, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia
Fax: +61 2 6125 0750
Tel: +61 2 6125 5061
E-mail: jianw@rsc.anu.edu.au

(Received 10 May 2007, revised 4 July 2007, accepted 11 July 2007)

doi:10.1111/j.1742-4658.2007.06022.x

There have been many approaches to solving problems associated with protein solubility. This article describes the application of directed evolution to improving the solubility of the C-terminal metal-binding domain of aminopeptidase P from *Escherichia coli*. During the course of experiments, the domain boundary and sequence were allowed to vary. It was found that extending the domain boundary resulted in aggregation with little improvement in solubility, whereas two changes to the sequence of the domain resulted in dramatic improvements in solubility. These latter changes occurred in the active site and abolished the ability of the protein to bind metals and hence catalyze its physiological reaction. The evidence presented here has led to the proposal that metals bind to the intact protein after it has folded and that the N-terminal domain is necessary to stabilize the structure of the protein so that it is capable of binding metals. The acid residues responsible for binding metals tend to repel one another – in the absence of the N-terminal domain, the C-terminal domain does not fold properly and forms inclusion bodies. Evolution of the C-terminal domain has removed the destabilizing effects of the metal ligands, but in so doing it has reduced the capacity of the domain to bind metals. In this case, directed evolution has identified active site residues that destabilize the domain structure.

The *Escherichia coli* aminopeptidase P (AMPP) is a protease with subunits that consist of two domains. Solution studies have shown that the activity of AMPP is manganese-dependent [1], and structural studies have shown that its active site contains two metals that are coordinated by residues from the C-terminal domain [2]. AMPP has a structure that is similar to that of prolidase and creatinase, but it is a tetramer, whereas both prolidase and creatinase are dimers [3]. Creatinase is a metal-independent enzyme that has an active site in a similar location to that of AMPP, whereas prolidase requires two metals that are coordinated to the protein via residues homologous to those found in AMPP. Methionine aminopeptidase is a monomeric protein that consists of a single domain that has structural similarity to the C-terminal domain of AMPP. Like prolidase, methionine aminopeptidase requires two metals that are coordinated via residues homologous to those of AMPP. These observations indicate that the C-terminal domain of AMPP, with its 'pita-bread' fold, is both stable and capable of being utilized for a number of catalytic functions. For this reason, we isolated the section of the AMPP gene that codes for the C-terminal domain and expressed it in *E. coli*. Surprisingly, this catalytic domain proved to be insoluble. Initially, it was thought that the change in solubility was due to the

Abbreviations
AMPP, *Escherichia coli* aminopeptidase P; DHFR, dihydrofolate reductase; TMP, trimethoprim.

 © 2007 The Authors Journal compilation © 2007 FEBS

exposure of hydrophobic residues that were covered in the intact protein. It was reasoned that the domain could be readily 'solubilized' using directed evolution. That is, the residues responsible for the insolubility of the domain could be altered using directed evolution so that soluble mutants could be obtained.

There are a several methods available for evolving a protein to make it more soluble. The method used in this work will be described briefly here; a more detailed account can be found elsewhere [4]. The method relies on the fact that dihydrofolate reductase (DHFR) is necessary for the survival of *E. coli*, and that low concentrations of DHFR inhibitors (typically at 2 $\mu g \cdot mL^{-1}$), such as trimethoprim (TMP), are lethal to the organism [4]. However, DHFR is an extremely soluble protein that can be easily expressed at much higher levels of TMP than the normally lethal doses. Overexpression of DHFR effectively renders *E. coli* TMP-resistant. Thus, if a target protein is expressed as a fusion protein with DHFR, its overexpression in soluble form will lead to TMP resistance. However, if the fusion construct is insoluble, *E. coli* will be susceptible to the inhibitor. In order to increase solubility, the target gene is mutated – using either error-prone PCR or DNA shuffling [5] – and the genes in the resulting mutant library are again fused to that of DHFR. The resulting mutant fusion proteins can again be expressed in *E. coli*, and TMP resistance can be monitored. The genes of mutants that confer increased TMP resistance are isolated and shuffled, and the new mutant library is monitored for increasingly higher levels of TMP resistance. After several rounds of evolution, the mutated genes of the target protein that confer TMP resistance are isolated and expressed to confirm that increased solubility has been evolved. It should be noted that this selection method does not prevent mutations that result in a loss of functional activity.

The object of this study was to increase the solubility of the C-terminal domain of AMPP, and in so doing to determine which residues are responsible for its poor solubility. Mutations were to be mapped onto the known structure so that possible reasons for poor solubility could be determined. Does aggregation of the AMPP C-terminal domain occur due to hydrophobic patches on the surface of the domain, or do specific residues destabilize the domain? These are the types of question that were to be addressed with the data that we obtained.

Results

In this study, consideration was given to the starting point of the AMPP C-terminal domain as well as its sequence. The location of the domain boundary was estimated by inspection of the structure, and this was compared with fragment lengths obtained experimentally. The experimental approach involved nuclease digestion of the AMPP gene (*pepP*). The gene fragments gave rise to a series of protein fragments that were examined for their solubility by fusing them to DHFR and monitoring the absence or presence of TMP resistance. Several different-length fragments were selected for further study. The genes for these fragments were isolated and shuffled to produce a mutant library, the members of which were then monitored for their ability to confer increased TMP resistance when fused to DHFR. The genes corresponding to resistant fragments were sequenced. At this stage, mutants of a single-length fragment were selected for a further round of shuffling. Two further rounds of shuffling were completed before a mutated fragment was selected for expression, purification, and characterization. At this stage, further refinement of the domain size was carried out. The locations of mutations that conferred increased solubility were noted.

Screening for the boundary of the C-terminal AMPP domain

N-terminal deletions of AMPP were generated by exonuclease III digestion of the *pepP* gene. A set of nested truncated *pepP* genes was fused to that of DHFR in the fusion vector pJWL1030folA and transformed into competent *E. coli* cells. Two libraries of about 10 000 clones were screened against two concentrations of TMP: 2 $\mu g \cdot mL^{-1}$ and 20 $\mu g \cdot mL^{-1}$. After 3–5 days of incubation at 37 °C, in comparison to plates without TMP, about 5% of the colonies with the truncated AMPP fragments appeared on the plates with 2 $\mu g \cdot mL^{-1}$ TMP, whereas none were visible on plates with 20 $\mu g \cdot mL^{-1}$ TMP. Thirty colonies were selected from the plate with 2 $\mu g \cdot mL^{-1}$ TMP. Plasmids were isolated, and the genes corresponding to the truncated AMPP were analyzed by restriction digestion and sequenced.

It was found that the deletions ranged in size from 201 bp to 636 bp. The predicted C–terminal boundary of AMPP corresponded to a deletion of 522 bp or 174 amino acids, as judged by an inspection of the AMPP crystal structure [2]. Most of the AMPP fragments that were selected from the agar plate were close in size to the C-terminal AMPP fragment predicted on the basis of the structure. Two genes for truncated fragments were isolated from the fusion vector and cloned into the expression vector pJWL1030. These two fragments, shown schematically in Fig. 1, corresponded to deletions of 157 amino acids (AMPP#2) and 212 amino

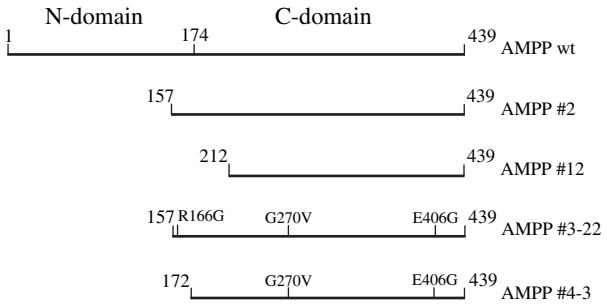

Fig. 1. Schematic diagram of AMPP. Wild-type AMPP consists of an N-terminal domain (1–174 amino acids) and a C-terminal domain (174–439 amino acids). C-terminal domain AMPP#2 has a 157 amino acid deletion, AMPP#12 has a 212 amino acid deletion, AMPP#3-22 has a 157 amino acid deletion, and AMPP#4-3 has a 172 amino acid deletion. Mutations are R166G, G270V, and E406G.

acids (AMPP#12). The truncated AMPP fragments were expressed and assayed for solubility, and neither gave rise to detectable levels of protein using the GelCode Blue stain reagent as detector, as shown in Fig. 2.

Improving solubility of the AMPP C-terminal domain

The first round of shuffling was screened with 5 $\mu g \cdot mL^{-1}$ TMP and utilized the genes of the five most common fragments found after screening for the domain boundary. These fragments correspond to deletions of 127, 143, 144, 157 and 212 amino acids, respectively. The DNA for the AMPP fragments was isolated from a number of resistant colonies and sequenced (Table 1). As can be seen, after the second round of DNA shuffling, all the chosen colonies gave fragments of the same length – all were derived from the AMPP#2 fragment (Fig. 1). Most of the mutant genes contained multiple mutations, two of which involved metal-binding ligands. The D271N and E406G mutations were expected to diminish or abolish the capacity of AMPP to bind metals. The results of subsequent rounds of evolution are also shown in Table 1. A number of mutations from round 1 disappeared in rounds 2 and 3, whereas the E406G mutation became common to all the mutants that were selected for sequencing. The G270V mutation appeared in the second round, and was found in all but one mutant protein selected in the third round. This latter mutation appeared to be incompatible with the D271N mutation; however, its close proximity to a metal-binding ligand suggested that it could (like the D271N mutation) also reduce or eliminate the capacity of the protein to bind metal. The

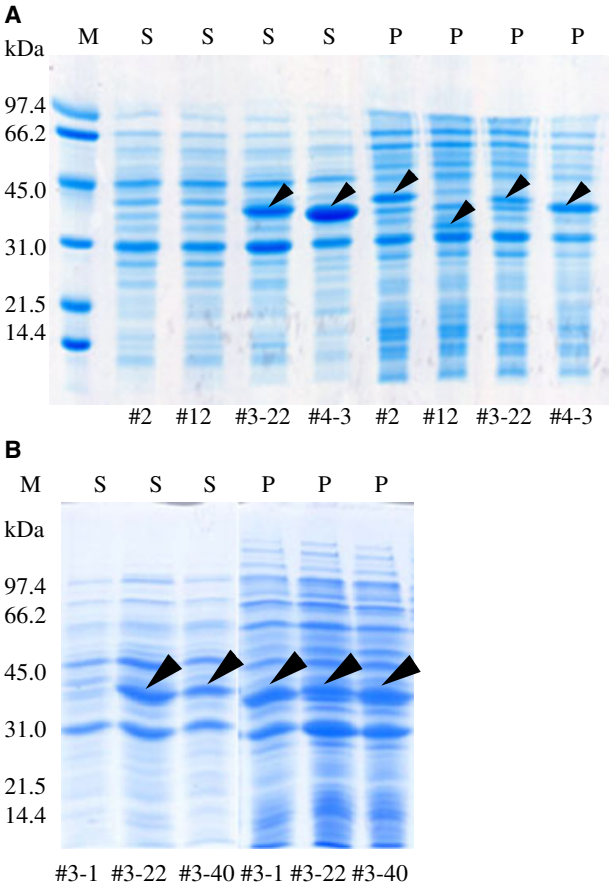

Fig. 2. Expression patterns of C-terminal AMPP domains. (A) An aliquot of supernatant (S) or pellet (P) from cells containing AMPP domains (#2, #12, #3-22, or #4-3) was denatured and resolved by 15% SDS/PAGE. (B) An aliquot of supernatant (S) or pellet (P) from cells containing AMPP domains (#3-1, #3-22, or #3-40) was denatured and resolved by 15% SDS/PAGE. Overexpressed AMPP domains are indicated by arrowheads. Low-range molecular mass standards (M) from Bio-Rad.

R166G mutation appeared in the first round of selection, increased in number in the second round, and was present in all but one of the round 3 mutant proteins. This mutation is close to the N-terminus of the fragment – it lies between the start of the fragment and the predicted start of the domain (Fig. 1). From the round 3 mutants, three were selected for further characterization: AMPP#3-1, AMPP#3-22, and AMPP#3-40. These fragments were subcloned so that they could be expressed without DHFR. The AMPP#3-22 mutant was clearly the most soluble (Fig. 2) and was chosen for further study. It is likely that the reduced solubility of the AMPP#3-40 mutant was due to the absence of the R166G mutation, whereas the reduced solubility of the AMPP#3-1 mutant could be attributed to a number of changes (Table 1).

Table 1. Sequence analysis of AMPP C-terminal domain mutants. The percentage of mutants containing a given mutation in each round is indicated.

Domains(deletion)	Mutations													
#1-1(157 aa)														
#1-9(157 aa)	R166G													
#1-21(157 aa)		V169A	E171G					D271N				E406G	D407N	V424M
#1-33(143 aa)				Y209H	H217R				V326I	P346L				
#1-40(157 aa)						C263Y						E406G		
% R1	20	20	20	20	20	20		20	20	20		40	20	20
#2-1(157 aa)				Y209H				D271N		P346L	P376L	E406G		
#2-5(157 aa)	R166G							D271N				E406G		
#2-6(157 aa)		V169A	E171G				G270V					E406G		
#2-13(157 aa)								D271N				E406G		
#2-30(157 aa)	R166G							D271N				E406G		
% R2	40	20	20				20	80		20	20	100		
#3-15(157 aa)	R166G	V169A	E171G					D271N				E406G		
#3-6(157 aa)	R166G						G270V					E406G		
#3-8(157 aa)	R166G						G270V					E406G		
#3-10(157 aa)	R166G						G270V					E406G		
#3-15(157 aa)	R166G						G270V					E406G		
#3-20(157 aa)	R166G						G270V					E406G		
#3-22(157 aa)	R166G						G270V					E406G		
#3-30(157 aa)	R166G						G270V					E406G		
#3-37(157 aa)	R166G						G270V					E406G		
#3-40(157 aa)					Y226C		G270V					E406G		
% R3	90	10	10		10		90	10				100		

The AMPP#3-22 mutant has the three most common mutations found in round 3: R166G, G270V, and E406G. The fragment was purified using two chromatographic steps, Q-SepharoseHP and SOURCE 15PHE. The purified fragment was then loaded onto a size exclusion column, and eluted in two peaks that corresponded to a monomer and a dimer of the fragment (Table 2). The fragment and the wild-type proteins were tested for enzymatic activity – only the wild-type protein displayed activity. Consistent with this lack of activity, atomic absorption measurements of the AMPP#3-22 mutant (as purified) gave no detectable trace of metals, demonstrating the inability of this mutant to bind metal ions. Furthermore, prolonged exposure of this fragment to high concentrations of divalent metal ions followed by dialysis to remove excess metal ions gave preparations of AMPP#3-22 that contain at most 0.15 ions per binuclear active site. This observation also argues for a very low binding affinity of the mutant fragment for metal ions. The residual metal ions (≤ 0.15) are adventitiously bound, as observed, for example, in other binuclear metalloenzymes, such as purple acid phosphatases and methionyl aminopeptidases [6–8].

Table 2. Size exclusion chromatography of AMPP C-terminal domains.

	Peak I (excluded)	Peak II (dimer)	Peak III (monomer)
AMPP#2 (refolded)	> 99%	–	–
AMPP#3-22	–	28%	72%
AMPP#4-3	–	–	> 99%

In vitro refolding

Wild-type AMPP and AMPP#3-22 were overexpressed and purified. Subsequently, the purified proteins were denatured with 6 M guanidine hydrochloride and renatured by dialysis in the presence of EDTA or metals, as described in Experimental procedures. Aggregated proteins were removed by centrifugation, and the proteins in the supernatant were analyzed by SDS/PAGE electrophoresis. The AMPP#2 fragment was expressed as an inclusion body and dissolved in 6 M guanidine hydrochloride. The denaturant was removed in the presence of EDTA or metals, and the soluble proteins were subjected to SDS/PAGE analysis. The results of these *in vitro* refolding attempts are shown in Fig. 3.

A previous study has shown that $ZnCl_2$ inhibits the activity of AMPP [1]. Here, the presence of $ZnCl_2$ in the dialysis buffer led to the precipitation of each of the three proteins. Neither the intact protein nor the

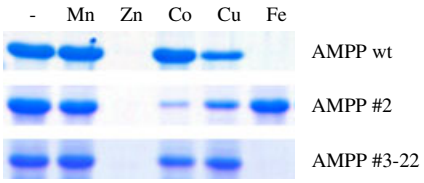

Fig. 3. *In vitro* refolding of AMPP and its C-terminal domains. Full-length AMPP (wt) and C-terminal domains (#2, #3-22) were denatured with 6 M guanidine hydrochloride and dialyzed overnight at 4 °C against 20 mM Tris (pH 7.6), containing 1 mM EDTA (–) or 1 mM various metals ($MnCl_2$, $ZnCl_2$, $CoCl_2$, $CuCl_2$ or $FeCl_3$). The precipitate was removed by centrifugation, and soluble proteins were resolved on a 15% SDS/PAGE gel.

fragments required metals to produce soluble protein. The wild-type and AMPP#3-22 proteins responded in a similar (although not identical manner) to the various metals. This observation, combined with the fact that AMPP#3-22 did not appear to bind metals, suggested that metals were not required for folding of the native enzyme or the AMPP#3-22 fragment. The response of the AMPP#2 fragment to metals differs from that of the wild-type protein or the AMPP#3-22 fragment. In order to investigate this difference further, the soluble AMPP#2 fragment (refolded with EDTA or metals) was loaded onto a size exclusion column. The fragment was excluded from the resin pores, suggesting that it had formed soluble microaggregates of partially unfolded protein (Table 2).

Evolution of the AMPP#3-22 fragment – optimizing the starting point

Exonuclease III digestion of the DNA corresponding to the AMPP#3-22 fragment was used to generate a library of N-terminal deletions of the fragment. This library was screened with a higher concentration of TMP than had been used in previous rounds of evolution. Several colonies were found to be resistant to 200 $\mu g \cdot mL^{-1}$ TMP. One of these colonies produced a fragment designated AMPP#4-3. DNA sequencing revealed that the size of the AMPP#4-3 fragment corresponded to a deletion of 172 amino acids from the wild-type sequence – this was very close to the boundary position predicted from an inspection of the structure. The DNA for this fragment was isolated from the fusion vector and cloned into the expression vector pJWL1030. The AMPP#4-3 fragment was expressed and assayed for solubility. From an inspection of Fig. 2, it appeared that *E. coli* produced more soluble AMPP#4-3 than AMPP#3-22. Whether AMPP#4-3 was more soluble than AMPP#3-22 was difficult to ascertain from the gel shown in Fig. 2, as there were background bands overlapping with that of the AMPP#4-3 fragment. To address this question of solubility, cells expressing AMPP#3-22 and AMPP#4-3 were grown on plates that contained TMP levels that ranged from 20 to 200 $\mu g \cdot mL^{-1}$. Both lines grew well on all the plates, suggesting that the solubility of the two fragments was similar. To ascertain the aggregation state of the AMPP#4-3 fragment, it was purified and analyzed by size exclusion chromatography. Unlike AMPP#3-22, AMPP#4-3 behaved as a monomer (Table 2), with no dimer component evident.

Discussion

Two approaches were taken to produce a soluble C-terminal domain of AMPP. Different-length domains were tested, and mutations were made to the sequences of these domains. It is known that the location of domain boundaries is critical to the formation of stable, correctly folded, isolated domains [9,10]. Domain boundaries can be predicted using sequence alignments or bioinformatic tools [11–14]. In the case of AMPP, a high-resolution structure is available, and it gives a good indication of where the C-terminal domain starts [2]. However, the expression of this domain based on the predicted boundary resulted in the production of inclusion bodies. This is not an uncommon problem, as noted by Holland *et al.* [15] – partitioning protein structure into domains is not always easy and successful. Two experimental approaches were considered as a means of correctly locating the domain boundary. First, consideration was given to limited proteolysis coupled with amino acid sequencing and MS [16,17]. Second, gene truncation has also been been used to obtain the soluble domains of multidomain proteins [18] – it is this method that was chosen for further study. This latter approach requires the construction of a truncation library and a method to screen for soluble domains [19].

A library of nested N-terminal deletions of the AMPP gene was created by exonuclease III digestion and subsequent screening by fusing them to the DHFR reporter gene and selecting with TMP. The initial round of truncations gave a series of deletions that allowed cells to survive on a minimal level of TMP. These domains were shuffled and one, AMPP#2, could be combined with mutations to produce a soluble domain. The AMPP#2 fragment was expressed, but gave rise to inclusion bodies – no soluble protein was detected. The fragment could be denatured, and it remained soluble upon removal of the denaturant. A sizing column revealed that the soluble form of the fragment consisted of a very high molecular mass

aggregate (> 200 kDa). Soluble variants of this fragment could be expressed in *E. coli* if suitable mutations were made to the DNA coding for AMPP#2. One of these variants, AMPP#3-22, was chosen for further study. Analysis with size exclusion chromatography revealed that AMPP#3-22 is a mixture of monomers and dimers. Only three mutations (R166G, G270V, and E406G) were required to convert the aggregated AMPP#2 fragment into the soluble AMPP#3-22 fragment. The first mutation (R166G) was removed in the final round of mutations in which the fragment length was varied to give the AMPP#4-3 fragment. This final fragment ran as a monomer when applied to a sizing column. This observation implicated the N-terminal peptide and the R166G mutation in the monomer–dimer equilibrium of AMPP#3-22. The AMPP#4-3 fragment has a length very close to that predicted for the C-terminal domain, on the basis of an inspection of the crystal structure (see above). Its amino acid sequence differs from that of the corresponding wild-type sequence at only two locations: positions 270 and 406. As noted in the previous section, E406 is a metal ligand that coordinates both metals, whereas G270 is adjacent to D271, which also coordinates both metals. The G270V and E406G mutations are likely to be responsible for the inability of the AMPP#3-22 fragment to bind metals. From these results, it appears that the solubility of the AMPP#4-3 fragment – or at least the ability to express this fragment in a soluble form – is connected with its inability to bind metals.

Metalloproteins can fold via metal-dependent or metal-independent pathways [20,21]. They may bind metal ions before polypeptide folding, after complete protein folding, or after partial folding. Phosphomannose isomerase is an example of a protein that requires a metal to fold. It requires zinc ions for both *in vivo* and *in vitro* folding [22]. The *in vitro* folding studies described in this article suggest that AMPP and C-terminal fragments fold in a metal-independent manner. Denatured AMPP and AMPP#3-22 both fold in the presence of EDTA, and both show similar folding patterns when exposed to metals during renaturation (Fig. 3). A plausible explanation for these observations is that the protein must be folded before metals bind – the metal-binding ligands must be appropriately placed to coordinate the incoming metals. Four acid residues coordinate the two divalent metal ions in the active site of AMPP (Fig. 4). The positively charged metals will neutralize the negatively charged acids. In the absence of metals, the negatively charged residues will tend to repel one another, thus destabilizing the protein. For the native protein, the presence of the N-terminal domain and the oligomeric structure of

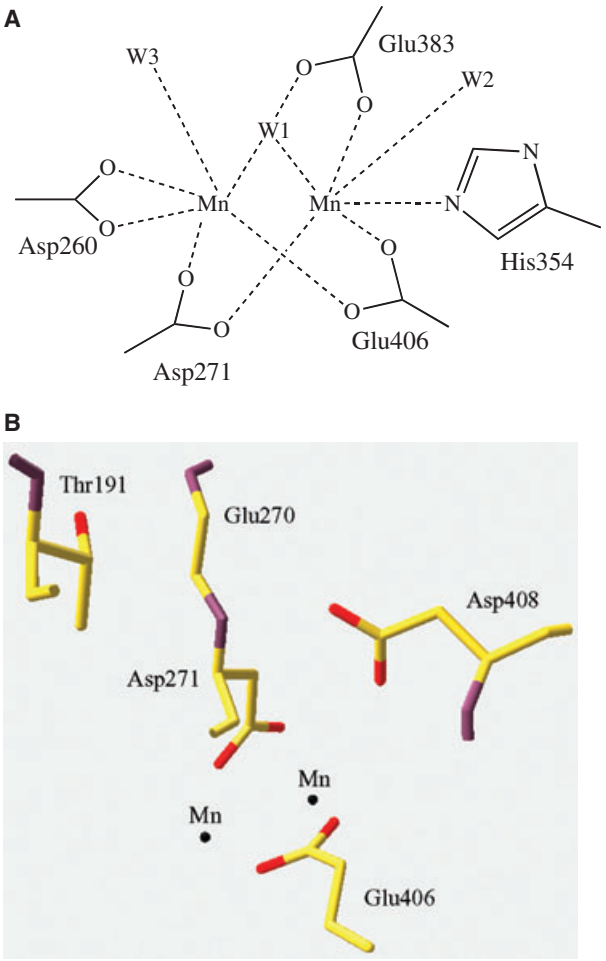

Fig. 4. The active site of AMPP. (A) Schematic diagram of the AMPP metal-binding sites. Metal-binding ligands are Asp260, Asp271, His354, Glu383, and Glu406. (B) Stereo view of the AMPP active site. Two mutations (Glu270 and Glu406) are responsible for improving the solubility of the C-terminal domain. The figure was generated from published data [27].

the protein may be necessary to maintain the structure of the C-terminal domain in a conformation that allows the metals to bind. Removing the N-terminal domain results in a C-terminal domain in which the acid residues of the active site repel one another, causing the protein to unfold (or to partially unfold). It is this unfolded form of the protein that aggregates and precipitates [23]. Mutations that abolish metal binding allow the peptide to assume a conformation close to that of the native protein – a stable conformation that results in soluble fragments that are incapable of binding metals.

The two rounds of evolution to optimize the starting point of the AMPP domain had opposing effects – the first round extended the domain size, whereas the last

round moved the starting point close to that predicted on the basis of an inspection of the structure. It would appear that extending the domain boundary had the effect of producing a slightly soluble aggregated form of the protein. Subsequent changes to the amino acid sequence were far more effective in improving the solubility of the domain. In the case of the AMPP protein, the boundary of the domain would have been better determined from an inspection of the structure rather than by the experimental methods that were used. The reasons for this are related to the metal-binding properties of the domain, and these will not necessarily affect studies with many other proteins. In the case of a stable, soluble domain, the methods described in this article should prove effective in locating the starting point of the domain.

In summary, directed evolution has been used to address the question of what causes the insolubility of the C-terminal domain of AMPP. The answer is relatively simple – modifying two active site residues can produce a soluble fragment. The E406G mutation converts a metal-binding ligand to a residue that is unlikely to bind metal. The G270V residue is located next to a metal-binding residue – this mutation is likely to cause a conformational change that is likely to further reduce the capacity of the fragment to bind metals. The conformational change could move E271 away from the active site, hence stabilizing the structure of the domain. In agreement with this interpretation, metal ion analysis of AMPP#3-22 by atomic absorption spectroscopy demonstrates that this mutant fragment has abolished the ability to bind metal ions. Although these two mutations dominate the list of mutations in round 3, it should be clear from the earlier round of shuffling that the mutation rate is considerably higher than two changes per round. Given the size of the mutant libraries (150 000), it is evident that the effects of all other mutations are significantly smaller than those of E406G and G270V. This idea is supported by the data shown in Table 1. By round 3, most of the mutations found in round 1 have been lost. Normally, one would expect an increase in the number of mutations per gene; however, we observed a decrease in the number of mutations per gene. The implication of this observation is that the effects of most mutations are small compared with those of G270V, E406G, and R166G. Changes at the surface of the protein do not appear to be major contributors to the solubility of the AMPP fragments. The AMPP protein appears to have evolved so that the metal-binding ligands are positioned optimally for the coordination of incoming metals. Metal binding would therefore stabilize the structure. One would expect that proteolysis could be used to produce stable C-terminal fragments, as these experiments could be conducted once metals have been bound. However, fragments identified in this manner may not fold when expressed in *E. coli*. The results presented in this article may explain the size of AMPP. It is a noncooperative tetramer that is considerably larger than, for example, the monomeric single-domain AMPM protein [3]. In the case of AMPP, the N-terminal domain appears to have a function in protein folding. Clearly, the single-domain AMPM protein has found another solution to this problem.

Experimental procedures

Chemicals and bacterial strains

All chemicals were purchased from Sigma-Aldrich (St Louis, MO). Molecular biology reagents and enzyme were brought from Roche (Basel, Switerland), New England Biolabs (La Jolla, CA), Bio-Rad (Hercules, CA), Novagen (Kilsyth, Australia), or GE Healthcare (Chalfont St Giles, UK). Primers were obtained from GeneWork (Thebarton, Australia). DNA purification kits (Qiagen, Doncaster, Australia) were used for all DNA isolations and purifications.

The *E. coli* strain DH5α (*sup*E44Δ*lac*U169 ø80 *lac*ZΔM15 *hsd*R17 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *rel*A1) was used for all aspects of the work. Cells were grown at 37 °C. Cell lines were maintained on LB medium agar plates supplemented with 50 $\mu g \cdot mL^{-1}$ kanamycin to maintain plasmids expressing recombinant *E. coli* AMPP and its domain variants.

Creating a library for truncated AMPP fragments

The 1.3 kb *pepP* gene encoding *E. coli* AMPP was PCR amplified from plasmid pPL670 [2] using a forward primer (5′-CCAAGCTTGTCGACGATGAGTGAGATATCC CGG-3′) and a reverse primer (5′-CGGGAATTCCTG CAGTTGCTTTCTCGCAGCAAC-3′), and then cloned between the *Sal*I and *Pst*I sites of the DHFR fusion vector pJWL1030folA [4] to produce pJWL1030folA–pepP. N-terminal deletions of AMPP were generated by partially digesting the *pepP* gene with exonuclease III in a manner similar to that described by Henikoff [24] and Ostermeier *et al.* [25]. pJWL1030folA–pepP (1–5 μg) was cut (linearized) at the 5′-end of *pepP* with *Sal*I. The *Sal*I-digested pJWL1030folA–pepP was digested with exonuclease III for varying times to generate nested deletions [25]. The truncated *pepP* fragments were then treated with Mung Bean Nuclease to remove single-strand DNA tails, and Klenow fragment DNA polymerase I was added to flush the DNA ends. The truncated DNA fragments were released from the pJWL1030folA vector by *Pst*I digestion, and subsequently separated on an agarose gel. The *pepP* fragments

with sizes between 0.9 kb and 1.3 kb were purified from the agarose gel. The DHFR fusion vector pJWL1030folA was digested with *Sal*I, and then incubated with Klenow fragment DNA polymerase I to produce blunt ends. The vector was further digested with *Pst*I. The truncated *pepP* fragments were then ligated to the blunt end and *Pst*I site of pJWL1030folA. Finally, the ligation mixture was transformed into DH5α cells by electroporation.

DNA shuffling

Random mutations were introduced into the *pepP* gene using DNA shuffling as described by Stemmer [26]. The shuffled *pepP* genes were ligated between the *Nde*I and *Pst*I sites of pJWL1030folA. The plasmid was then transformed into cells by electroporation.

Selection for TMP resistance

The truncated *pepP* gene library was plated on Mueller–Hinton agar (Difco, Becton Dickinson, Sparks, MD) plates that were supplemented with 50 $\mu g \cdot mL^{-1}$ kanamycin and 2 or 20 $\mu g \cdot mL^{-1}$ TMP. The TMP-resistant colonies appeared after incubation at 37 °C for 3–5 days.

The transformed cells with shuffled *pepP* genes were plated on the Mueller–Hinton agar plates supplemented with 50 $\mu g \cdot mL^{-1}$ kanamycin and increasing concentrations of TMP for the three rounds of evolution. For the first round, 5 $\mu g \cdot mL^{-1}$ TMP was used, and in the second and third rounds, 10 and 20 $\mu g \cdot mL^{-1}$ TMP were used, respectively. In each round, a library of 150 000 colonies was screened. The DNA for the 10 mutant genes from round 1 was shuffled for selection in round 2, and 18 genes were selected from round 2 and shuffled for selection in round 3.

Protein expression and solubility assay

The intact AMPP as well as the C-terminal fragments of AMPP were expressed in the same manner. The genes were PCR amplified and cloned between the *Nde*I and *EcoR*I sites of the pJWL1030 expression vector [4]. The plasmids were then transformed into cells by electroporation. Cells expressing each of these domains were grown overnight at 4 °C in LB medium containing 50 $\mu g \cdot mL^{-1}$ kanamycin. Cells were harvested and lysed using the BugBuster detergent (Novagen). Solubility assays were carried out using SDS/PAGE gel electrophoresis and staining using the GelCode Blue stain reagent (Pierce, Rockford, IL) as described elsewhere [4].

Protein purification and activity assay

The wild-type AMPP as well as C-terminal domains of AMPP were purified using a modified form of the protocol used for AMPP [2]. Briefly, cells were harvested and resuspended in 20 mM Tris (pH 7.6), and then lysed using a French press. The lysates were centrifuged at 30 000 ***g*** for 40 min at 4 °C (Sorvall RC5C, Thermo Electron, with SS34 rotor), and the supernatants were applied to a Q-SepharoseHP column (GM Healthcare) and eluted with a gradient of 0–1 M NaCl in 20 mM Tris (pH 7.6). Pooled fractions were combined with an equal volume of 20 mM Tris (pH 7.6) and 3 M $(NH_4)_2SO_4$. After centrifugation as above, the supernatant was applied to a SOURCE 15PHE column (GE Healthcare) and eluted with a gradient of 1.5–0 M $(NH_4)_2SO_4$ in 20 mM Tris (pH 7.6). The pooled fractions were dialyzed against 20 mM Tris (pH 7.6), and concentrated using Centriplus filter devices (YM-10; Millipore, Bedford, MA). The enzymatic activities of intact and C-terminal domains of AMPP were assayed using the quenched fluorescent substrate Lys(Abz)-Pro-Pro-pNA (Bachem, Bubendorf, Switzerland), as described elsewhere [27].

In vitro refolding

The purified AMPP (wild-type) and AMPP#3-22 were denatured with 6 M guanidine hydrochloride in the presence of 1 mM EDTA or 1 mM various metals ($MnCl_2$, $ZnCl_2$, $CoCl_2$, $CuCl_2$, or $FeCl_3$). The denatured proteins were dialyzed at 4 °C overnight against 20 mM Tris (pH 7.6) with EDTA or metals. The inclusion bodies formed from AMPP#2 were dissolved in 6 M guanidine hydrochloride, and then dialyzed against 20 mM Tris (pH 7.6) with EDTA or metals. After dialysis, the solutions containing AMPP, AMPP#2 and AMPP#3-22 were centrifuged at 16 000 ***g*** for 10 min at 4 °C (Sorvall RC5C with SS34). The supernatants and pellets were separated. The pellets were mixed with 20 mM Tris (pH 7.6) and vortexed to ensure that they were resuspended. Equal volumes of the solutions containing the supernatants and the resuspended pellets were run on a 15% SDS/PAGE gel and stained using the GelCode Blue stain reagent.

Size exclusion chromatography

A gel filtration assay was carried out using a Superdex 200 HP 10/30 column (GM Healthcare). The column was equilibrated with 20 mM Tris (pH 7.6) and 0.15 M NaCl, and calibrated with a marker mix including aldolase (158 kDa, GM Healthcare), phosphotriesterase (74 kDa) [28] and dienelactone hydrolase (26 kDa) [29].

Metal ion analysis

Metal ion concentrations were determined in triplicate by atomic absorption spectroscopy using a Varian SpectrAA 220FS instrument. Standard solutions for Fe^{2+}, Mn^{2+},

Zn^{2+} and Co^{2+} ranged from 20 p.p.b. to 200 p.p.b., and were prepared from analytical stock solutions (Merck, Kilsyth, Australia) using MilliQ water (produced by MilliQ reagent water system; Millipore). Aliquots of purified protein samples were sufficiently diluted with MilliQ to obtain metal ion concentrations in the range between 20 p.p.b. and 200 p.p.b., assuming a full complement of two metals per active site. The quantity of metal ions in MilliQ water was below the detection limit of the instrument. The estimated error for each measurement was less than 5%.

Acknowledgements

The authors thank Cameron McRae of the Bimolecular Resource Facility for DNA sequencing, and Professor Nick Dixon for providing plasmid pPL670.

References

1 Graham SC, Bond CS, Freeman HC & Guss JM (2005) Structural and functional implications of metal ion selection in aminopeptidase P, a metalloprotease with a dinuclear metal center. *Biochemistry* **44**, 13820–13836.
2 Wilce MC, Bond CS, Dixon NE, Freeman HC, Guss JM, Lilley PE & Wilce JA (1998) Structure and mechanism of a proline-specific aminopeptidase from Escherichia coli. *Proc Natl Acad Sci USA* **95**, 3472–3477.
3 Bazan JF, Weaver LH, Roderick SL, Huber R & Matthews BW (1994) Sequence and structure comparison suggest that methionine aminopeptidase, prolidase, aminopeptidase P, and creatinase share a common fold. *Proc Natl Acad Sci USA* **91**, 2473–2477.
4 Liu JW, Boucher Y, Stokes HW & Ollis DL (2006) Improving protein solubility: the use of the Escherichia coli dihydrofolate reductase gene as a fusion reporter. *Protein Expr Purif* **47**, 258–263.
5 Neylon C (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res* **32**, 1448–1459.
6 Schenk G, Boutchard CL, Carrington LE, Noble CJ, Moubaraki B, Murray KS, de Jersey J, Hanson GR & Hamilton S (2001) A purple acid phosphatase from sweet potato contains an antiferromagnetically coupled binuclear Fe–Mn center. *J Biol Chem* **276**, 19084–19088.
7 Larrabee JA, Leung CH, Moore RL, Thamrong-Nawasawat T & Wessler BS (2004) Magnetic circular dichroism and cobalt(II) binding equilibrium studies of Escherichia coli methionyl aminopeptidase. *J Am Chem Soc* **126**, 12316–12324.
8 Mitic N, Smith SJ, Neves A, Guddat LW, Gahan LR & Schenk G (2006) The catalytic mechanisms of binuclear metallohydrolases. *Chem Rev* **106**, 3338–3363.
9 Xu Y, Wen D, Clancy P, Carr PD, Ollis DL & Vasudevan SG (2004) Expression, purification, crystallization, and preliminary X-ray analysis of the N-terminal domain of Escherichia coli adenylyl transferase. *Protein Expr Purif* **34**, 142–146.
10 Kerr ID, Berridge G, Linton KJ, Higgins CF & Callaghan R (2003) Definition of the domain boundaries is critical to the expression of the nucleotide-binding domains of P-glycoprotein. *Eur Biophys J* **32**, 644–654.
11 Rigden DJ (2002) Use of covariance analysis for the prediction of structural domain boundaries from multiple protein sequence alignments. *Protein Eng* **15**, 65–77.
12 Dumontier M, Yao R, Feldman HJ & Hogue CW (2005) Armadillo: domain boundary prediction by amino acid composition. *J Mol Biol* **350**, 1061–1073.
13 Liu J & Rost B (2004) Sequence-based prediction of protein domains. *Nucleic Acids Res* **32**, 3522–3530.
14 Galzitskaya OV & Melnik BS (2003) Prediction of protein domain boundaries from sequence alone. *Protein Sci* **12**, 696–701.
15 Holland TA, Veretnik S, Shindyalov IN & Bourne PE (2006) Partitioning protein structures into domains: why is it so difficult? *J Mol Biol* **361**, 562–590.
16 Severinova E, Severinov K, Fenyo D, Marr M, Brody EN, Roberts JW, Chait BT & Darst SA (1996) Domain organization of the Escherichia coli RNA polymerase sigma 70 subunit. *J Mol Biol* **263**, 637–647.
17 Christ D & Winter G (2006) Identification of protein domains by shotgun proteolysis. *J Mol Biol* **358**, 364–371.
18 Hart DJ & Tarendeau F (2006) Combinatorial library approaches for improving soluble protein expression in Escherichia coli. *Acta Crystallogr D Biol Crystallogr* **62**, 19–26.
19 Cornvik T, Dahlroth SL, Magnusdottir A, Flodin S, Engvall B, Lieu V, Ekberg M & Nordlund P (2006) An efficient and generic strategy for producing soluble human proteins and domains in E. coli by screening construct libraries. *Proteins* **65**, 266–273.
20 Wittung-Stafshede P (2004) Role of cofactors in folding of the blue-copper protein azurin. *Inorg Chem* **43**, 7926–7933.
21 Wilson CJ, Apiyo D & Wittung-Stafshede P (2004) Role of cofactors in metalloprotein folding. *Q Rev Biophys* **37**, 285–314.
22 Proudfoot AE, Goffin L, Payton MA, Wells TN & Bernard AR (1996) In vivo and in vitro folding of a recombinant metalloenzyme, phosphomannose isomerase. *Biochem J* **318** (2), 437–442.
23 Villaverde A & Carrio MM (2003) Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol Lett* **25**, 1385–1395.

24 Henikoff S (1987) Unidirectional digestion with exonuclease III in DNA sequence analysis. *Methods Enzymol* **155**, 156–165.
25 Ostermeier M, Nixon AE, Shim JH & Benkovic SJ (1999) Combinatorial protein engineering by incremental truncation. *Proc Natl Acad Sci USA* **96**, 3562–3567.
26 Stemmer WP (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci USA* **91**, 10747–10751.
27 Graham SC, Lilley PE, Lee M, Schaeffer PM, Kralicek AV, Dixon NE & Guss JM (2006) Kinetic and crystallographic analysis of mutant Escherichia coli aminopeptidase P: insights into substrate recognition and the mechanism of catalysis. *Biochemistry* **45**, 964–975.
28 Yang H, Carr PD, McLoughlin SY, Liu JW, Horne I, Qiu X, Jeffries CM, Russell RJ, Oakeshott JG & Ollis DL (2003) Evolution of an organophosphate-degrading enzyme: a comparison of natural and directed evolution. *Protein Eng* **16**, 135–145.
29 Kim HK, Liu JW, Carr PD & Ollis DL (2005) Following directed evolution with crystallography: structural changes observed in changing the substrate specificity of dienelactone hydrolase. *Acta Crystallogr D Biol Crystallogr* **61**, 920–931.