

Manuscript EMBO-2014-89367

Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures

Sebastian Zoll, Stancho Stanchev, Jakub Began, Jan Skerle, Martin Lepšík, Lucie Peclinovska, Pavel Majer and Kvido Strisovsky

Corresponding author: Kvido Strisovsky, Academy of Sciences of the Czech Republic

Review timeline:

Submission date:	25 June 2014
Editorial Decision:	15 July 2014
Revision received:	08 August 2014
Accepted:	19 August 2014

Editor: Hartmut Vodermaier

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision

15 July 2014

Thank you again for submitting your manuscript for consideration by the EMBO Journal. It has now been seen by three referees, whose comments are shown below. I am pleased to communicate that all three referees are quite enthusiastic about this work, and we shall therefore be happy to offer publication in our journal, pending adequate revision of a number of (mostly presentational) issues. In this respect, please note that I do not share referee 2's concern regarding the homology model in Figure 6B, which I think makes a nice contribution and is appropriately placed in the discussion section; but it may be helpful to emphasize the speculative nature of the model a bit more in the text.

When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Review Process File, and will therefore be available online to the community. Please also be reminded that it is EMBO Journal policy to allow only a single round of revision, and acceptance of your manuscript will therefore depend on the completeness of your responses in this revised version. Finally, as a matter of editorial policy, competing manuscripts published during the revision period will not negatively impact on our final decision on your manuscript.

Thank you for the opportunity to consider your work for publication! I look forward to your revision.

Referee #1:

Review of "Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures"

This manuscript describes a comprehensive and elegant study of the interactions between peptide-based inhibitors and a bacterial rhomboid protease, GlpG. Though GlpG has previously been crystallized with inhibitors, the nature of the inhibitors used in this study provided a much more intimate view of likely enzyme-substrate interactions, extending far out of the site of actual peptide bond cleavage. Thorough sequence analyses of inhibitor activity and of enzyme activity on substrate variants accompany the structural studies and provide insight into the amino acid preferences at each substrate peptide position. In addition, the structure is analyzed to the fullest extent, both statically and using it as a basis for molecular dynamics. In the discussion of their findings, the authors engage seriously and thoughtfully with previous literature and models. The figures are clear and well-labeled.

This is one of the most engaging manuscripts I have ever read, presenting a huge amount of information (but distilled down to a few main messages) aesthetically, seamlessly, and convincingly. The findings will provide much food for thought and discussion in the large community of investigators interested in rhomboid proteases, particularly with regard to the roles of the L1 loop, TMD5 motions, and His150. More significantly, this study can be mined for its implications for substrate specificity in other rhomboids.

This referee absolutely and enthusiastically recommends publication.

A ridiculously trivial list of typos appears below, just because I may as well point them out.

p. 6 There is an extra period near the beginning of the last paragraph.

p. 8 In line 5, it should be "... towards Q189 (Fig. 3D), which has been..."

p. 9 There is an extra period in the 5th line from the bottom.

p. 11 Fourth line from the top, should it be "above the S1' and S2' subsites"?

p. 11 Fifth line from the top, there is an extra "and"

p. 11 Should be "This is in agreement with a recent report..."

p. 13 There is an extra comma after a period in the first line.

p. 13 One instance of TMD5 is written as TMD 5.

p. 13 Should be a comma to yield: "accompany ligand binding, and H150 is worth highlighting."

p. 13 Consider changing "... to make space for the P2 residue of the ligand, which can accommodate almost any residue (Fig. 2A)." It should perhaps instead be, "to make space for the P2 residue of the ligand, which can be almost any amino acid type (Fig. 2A)." The reasoning is that the enzyme would do the accommodating of any residue at P2, so it is not the P2 itself that accommodates.

Referee #2:

This manuscript describes the development of novel chloromethylketone peptidomimetic inhibitors of the *E. coli* rhomboid serine protease GlpG and the use of these inhibitors for cocrystallization and elucidation of the structural basis for substrate recognition. This is the first report of the cocrystallization of GlpG with peptide analogues that closely mimic substrate. Several important findings from this study include the identification of substrate-binding pockets S1-S4, the participation of loop 5 in the S4 pocket, analysis of residues putatively contributing to the oxyanion hole, and the conformational changes in specific side chains upon binding to the substrate-based inhibitor. There are, however, several important issues that require attention:

(1) The TatA substrate is said to be a natural rhomboid substrate. This is true, but not for the specific rhomboid investigated here (*E. coli* GlpG). This should be clarified wherever it is mentioned (e.g., the abstract).

(2) Page 5: the MS data showing covalent binding of inhibitor to the wild-type enzyme but not the H254A or S201A mutants does not mean that the inhibitor is directly bound to these residues. Proof of this comes later, with the determination of the crystal structures. This should be clarified here on

page 5.

(3) Page 6: Of the 5 peptide inhibitor variants tested, all include mutations that eliminate substrate cleavage. It would strengthen the case that these inhibitors truly behave like substrates to have one or two compounds in which mutation still allows cleavage of substrate and retains activity in the inhibitor.

(4) Page 8: Why is it surprising that substrates with small P4 residues were cleaved less efficiently with F146A and F146I GlpG mutants? As Ala and Ile side chains are smaller than that of Phe, the S4 site should be larger, making smaller P4 residues less preferable.

(5) Fig. 4b: For the bottom right panel (F146I mutant), why isn't W196 shown as in the other two cases? It is harder to compare these structures without showing the same residues.

(6) Page 10: Couldn't movement of F232 and W236 side chains be considered consistent with these residues participating in substrate gating? In any event, regarding substrate gating, it's important to emphasize here that the bound peptidomimetic does not include the prime side or the rest of the substrate TMD. Thus, these structures give information about the covalent enzyme-substrate intermediate, not how the entire substrate interacts with the enzyme and enters the active site.

(7) Page 10-11: It is difficult to determine the validity of the molecular dynamics results, as P1', P2' and P3' substrate positions are added only computationally.

(8) Page 12-13: The homology modeling of TatA TMD substrate and docking to GlpG is highly speculative and should be excluded. Why is this added only to the discussion section?

(9) Page 14: It is stated "Although the whole cavity forming the S1 subsite is quite large, only small residues, such as alanine, cysteine or serine are accepted in the P1 position of the substrate. One explanation for this could be that larger side chains could interfere with the stability of the water molecules in the water retention site and thus inhibit the deacylation step." But the CMK inhibitors with large P1 residues are also not accepted and these do not undergo deacylation in working as inhibitors. This seems to be an inconsistency.

Referee #3:

The manuscript by Zoll et al describes a series of structures obtained with the GlpG rhomboid protease from *E. coli* covalently bound to novel chloromethylketone-based inhibitors designed to mimic the substrate-bound acyl-enzyme intermediate. This represents a significant advance in the elucidation of rhomboid function, since up to this point there was no structural information available on substrate-enzyme interactions for any rhomboid. While a number of structures of GlpG in complex with small, mechanism-based inhibitors have been published, these are the first to reveal the structural basis for rhomboid substrate specificity.

To lay the groundwork for this study, it was necessary to determine the sequence specificity of *E. coli* GlpG for positions P1 to P4 of the substrate, expanding on previous work that had defined a general rhomboid substrate consensus sequence (Srisovsky et al, Mol Cell 2009). Tetrapeptide chloromethylketone inhibitors based on these sequences were shown to inhibit GlpG irreversibly, and followed predictions based on the substrate consensus sequence, providing good evidence that these interactions are the same as those between GlpG and its substrate. X-ray crystal structures of a chloromethylketone inhibitor covalently bound via the active site serine and histidine residues revealed new details regarding the role of loops 5 and 3 in binding the substrate, and the presence of a structurally malleable hydrophobic pocket formed by the structured L1 domain. Molecular dynamics simulations on a model of the Michaelis complex based on this crystal structure provided additional insight into potential interactions with the C-terminal region of the substrate (P1' to P3').

Overall, this is a well-executed study that provides several new insights for rhomboid proteases and how they interact with their substrate sequences. The role of the conserved but enigmatic L1 loop in substrate binding will be particularly interesting, not only for researchers working on enzymatically active members of the rhomboid family, but also those focused on inactive members (iRhoms) and the related derlins due to its potential role in binding target proteins.

Minor points to address:

Page 7 - The meaning of the following phrase is not clear, and should be reworded:
"Its part more distal to the P1 side chain, although buried in the core of the protein,..."

Page 14 - It is suggested that the reason the S1 pocket is not able to accommodate a larger side chain

is that it would interfere with the stability of water molecules in the water retention site, and therefore inhibit deacylation. However, this would suggest that an acyl-enzyme intermediate could be trapped, an outcome that should give rise to a new species in the gel-based assay used to screen this substrate library yet was not observed. This should be acknowledged, or this speculation removed.

Figure 4A - Error bars are implied in the figure legend, but not shown in the figure.

Supplemental Figure S7 - Panel C appears to have been omitted, although it is described in the text and figure legend.

1st Revision - authors' response

08 August 2014

Response to reviews

We thank the anonymous referees for their criticism and input. We have strived to address their comments; our responses are summarised below. The text passages that have been changed or added in the manuscript during the revision and are cited literally have been highlighted in blue. The revised manuscript also contains minor modifications throughout to correct the typos and increase clarity in some places.

Referee #1:

Referee #1 finds our manuscript very engaging, and „absolutely and enthusiastically recommends publication“, which is very pleasing to read. The reviewer points out a number of typos, for which we are grateful. We made sure that all of them have been corrected in the revised version of the manuscript.

Referee #2:

This referee considers our findings important, but finds several important issues that require attention. We address them all to full extent below.

Original comment:

(1) The TatA substrate is said to be a natural rhomboid substrate. This is true, but not for the specific rhomboid investigated here (E. coli GlpG). This should be clarified wherever it is mentioned (e.g. the abstract).

Response:

This is of course absolutely correct. To make this point even clearer we have corrected the respective sentence in the abstract to:

„Peptidyl-CMKs derived from the natural bacterial rhomboid substrate TatA from *Providencia stuartii* bind GlpG in a substrate-like manner, and...“

and in the first paragraph of the Discussion (p. 13) to „We have developed a new series of peptidic chloromethylketone inhibitors, based on a natural bacterial rhomboid substrate sequence (*Providencia stuartii* TatA) (Stevenson et al, 2007), that provide the first structural insight into substrate binding to rhomboids.“

Original comment:

(2) Page 5: the MS data showing covalent binding of inhibitor to the wild-type enzyme but not the H254A or S201A mutants does not mean that the inhibitor is directly bound to these residues. Proof of this comes later, with the determination of the crystal structures. This should be clarified here on page 5.

Response:

To accommodate this relevant comment we have now rephrased the respective sentences to:
„The compound Ac-IAAA-cmk inhibited GlpG in a concentration and time-dependent manner (Fig. 1B and S2A), and mass-spectrometric analysis indicated that it formed a stoichiometric (1:1) complex with the enzyme, which was dependent on the catalytic residues Ser201 and His254 (Fig.

S2B). Upon reaction of Ac-IAAA-cmk with wild type (wt) GlpG, but not with its S201A and H254A mutants, a faster migrating species on SDS-PAGE arose (Fig. 1B and S2B). A similar effect has been observed recently upon disulfide cross-linking of transmembrane domains (TMDs) 2 and 5 in GlpG (Xue & Ha, 2013), which suggested that Ac-IAAA-cmk may be crosslinking two TMDs of GlpG. The mass shift of GlpG in the presence of Ac-IAAA-cmk was consistent with the formation of the inhibitor-enzyme complex and elimination of a leaving group of approximately 36 Da (consistent with the molecular weight of HCl). This behaviour was analogous to how CMKs react with classical serine proteases, and we concluded that Ac-IAAA-cmk acted as a mechanism-based inhibitor of GlpG, forming a covalent adduct with the catalytic dyad residues, thus cross-linking TMDs 4 and 6. Furthermore, N-terminal truncation analysis of Ac-IAAA-cmk revealed that the inhibitory potency markedly decreased with progressive truncation of peptidyl chain of the inhibitor (Fig. 1C).“

We have also re-measured the mass spectra and provide higher quality data in Fig. S2B.

Original comment:

(3) Page 6: Of the 5 peptide inhibitor variants tested, all include mutations that eliminate substrate cleavage. It would strengthen the case that these inhibitors truly behave like substrates to have one or two compounds in which mutation still allows cleavage of substrate and retains activity in the inhibitor.

Response:

We thank this referee for this very relevant point. Although one such inhibitor had already been included in the paper in Figures 1C and S2A, this might not have been very obvious from the text. We have therefore taken the referee's point very seriously and generated another inhibitor that harbours **two** mutations that still allow substrate cleavage – A6S and T7K - to produce inhibitor Ac-ISKA-cmk. The inhibitory properties of this compound are indistinguishable from the parent compound Ac-IATA-cmk within experimental error. The inhibition data are shown in Fig. 2C, and the example source data in Fig. S4. In total, we now present 3 mutations that do not perturb the cleavage of substrate nor the potency of the respective inhibitors, and believe that we have thus strengthened the case.

We have updated the respective text in the manuscript as follows:

„While all the amino acid changes that impaired cleavage of mutant TatA substrates (I5S, I5G, A6D, A8V and A8G) also profoundly worsened the inhibitory properties of the variant peptidyl-CMKs, those amino acid changes that did not negatively affect cleavage of mutant substrate (T7A and A6S/T7K) had no impact on the inhibitory properties of the respective CMK derivatives (Fig. 1C, 2C, S4). This demonstrates that TatA-derived peptidyl-CMKs bind GlpG in a substrate-like manner and can hence be used as substrate mimetics in crystallographic experiments.“

Original comment:

(4) Page 8: Why is it surprising that substrates with small P4 residues were cleaved less efficiently with F146A and F146I GlpG mutants? As Ala and Ile side chains are smaller than that of Phe, the S4 site should be larger, making smaller P4 residues less preferable.

Response:

It is not surprising now, in retrospect, after we have solved the structures of the inhibitor complexes. What we actually meant to say was that it is surprising that the F146A mutant is active, unlike previously reported. We have now re-formulated this section to improve clarity:

„As P4 residue crucially contributes to substrate recognition by several rhomboids (Strisovsky et al, 2009), strongly influencing mainly the k_{cat} of the reaction (Dickey et al, 2013), we examined the functional and structural properties of S4 subsite in greater detail. The mutation of F146 to alanine was reported to inactivate GlpG without substantially affecting its thermodynamic stability (Baker & Urban, 2012), which was previously difficult to explain. Since F146 interacts with the P4 residue side chain of the substrate, we hypothesized that mutations in F146 could actually affect the P4 specificity of GlpG. To test this hypothesis we engineered complementary enzyme and substrate mutants by introducing hydrophobic residues of different side chain volumes to position 146 of GlpG (F146A and F146I) and by testing their activity against all 20 possible mutants in the P4 position of TatA substrate. Indeed, the F146A mutant was not inactive as previously reported (Baker & Urban, 2012), but it rather showed a shift in specificity for the P4 residue. TatA variants with

smaller residues in P4 position (e.g. A, C, V) were cleaved less efficiently by both the F146A and F146I mutants than by wt GlpG, while TatA variants with larger hydrophobic side chains in P4 position (such as M, F, W) were cleaved significantly better by F146A and F146I mutants than by wt GlpG (Fig. 4A, Fig. S7). “

Original comment:

(5) Fig. 4b: For the bottom right panel (F146I mutant), why isn't W196 shown as in the other two cases? It is harder to compare these structures without showing the same residues.

Response:

The bottom right panel of Fig. 4B shows only the amino acids of the enzyme that make contacts with the ligand in each case. Since W196 does not make contacts with the ligand in the F146I complex, it is not relevant and is not shown in the bottom right panel. If it were shown, it would have obscured F197, which does make contacts with the ligand and is thus relevant and shown in the panel. We believe that the upper panels allow comparison of the binding modes of the ligands in the structures, as they are oriented in the same way, and hope that the referee will appreciate our explanation.

Original comment:

(6) Page 10: Couldn't movement of F232 and W236 side chains be considered consistent with these residues participating in substrate gating? In any event, regarding substrate gating, it's important to emphasize here that the bound peptidomimetic does not include the prime side or the rest of the substrate TMD. Thus, these structures give information about the covalent enzyme-substrate intermediate, not how the entire substrate interacts with the enzyme and enters the active site.

Response:

We appreciate this very relevant comment. We are conscious of the fact that our ligands span only the non-prime side, are aware of the limitations of our crystal structures, and are confident that they nevertheless bring very useful implications for substrate access to rhomboid. We have made any comments about substrate access clearer, emphasising the relevant data and hypotheses and pointing out conceptual differences. We have taken care in clarifying and defining the term ‘gating’ and the ‘lateral gating’ hypothesis as it occurs in rhomboid literature. Our data are definitely consistent with F232 and W236 side chains participating with substrate binding to rhomboid, but that is not the same as ‘substrate gating’ or ‘lateral gating’. Not to confuse or amalgamate these distinct mechanisms, we emphasise that the ‘lateral gating’ model assumes a large shift in the position of TMD5 (in superficial analogy to the translocon structure), whereas our data, as well as data of others suggest that the binding of substrates into the active site of GlpG is feasible without any TMD5 displacement. The point is best illustrated by the citation of the respective re-written passage from p. 14-15 of the Discussion, which we hope will address the raised criticism.

“Structural changes in rhomboid accompanying substrate binding

Crystal structures of model intramembrane proteases suggest that substrate access to their catalytic residues may be conformationally regulated (Strisovsky, 2013). Based on the alternative conformation of one molecule in the asymmetric unit of a crystal structure of GlpG (Wu et al, 2006), substrate access to rhomboid protease had been suggested to be governed by a ‘gating’ mechanism. In analogy to the translocon (Van den Berg et al, 2004), this mechanism should involve a large dislocation of TMD5 to make the core of the enzyme accessible laterally from the lipid bilayer (Baker et al, 2007; Wu et al, 2006). Mutations in residue pairs W236A/F153A and F232A/W157A, designed to weaken the contacts between TMD2 and 5, increased enzymatic activity, supposedly by opening the TMD5 gate (Baker et al, 2007), which was further supported by subsequent enzymatic and thermodynamic studies (Baker & Urban, 2012; Moin & Urban, 2012). In contrast, other authors showed that preventing large lateral movement of TMD5 by chemically cross-linking TMDs 2 and 5 in a W236C/F153C mutant does not abrogate the activity of GlpG. This suggests that a ‘gating’ movement of TMD5 may not actually be required for substrate binding and it leaves the mechanism of substrate access to rhomboid controversial.

Our structures of the peptidyl-CMK complexes show that the L5 loop has to be displaced significantly to allow binding of substrate to the active site, but we do not observe any significant movement of the adjoining TMD5. Since our peptide ligands comprise only the non-prime side residues and capture the reaction at the stage of the tetrahedral or acylenzyme intermediate, we explored rhomboid-substrate interactions at the prime side and possible involvement of TMD5 by

molecular modelling and dynamics. The results show that a large lateral movement of TMD5 is not required for the formation of the acylenzyme nor the Michaelis complex with the P4 to P3' segment of the substrate. Our data are thus compatible with the published crosslinking data suggesting that major movements of TMD5 are not required for substrate access (Xue & Ha, 2013). We cannot formally exclude the possibility of a large TMD5 movement in the earlier phases of a transmembrane substrate binding. However, the positions of residues W236 and F153, which we observe in the Michaelis complex model (Fig. 6A and S8E), suggest that they may directly interact with the substrate, rather than just acting as 'openers' of the TMD5 gate. These results collectively imply that the lateral gate opening analogy with the translocon (Baker et al, 2007; Wu et al, 2006) may not be entirely correct and that substrate access mechanism to rhomboid merits further investigation."

Original comment:

(7) Page 10-11: It is difficult to determine the validity of the molecular dynamics results, as P1', P2' and P3' substrate positions are added only computationally.

Several previous papers in the field that presented 'models' of substrate bound in the active site that were either pure manual docking drawings or were based on small molecule inhibitor complex structures (Lemieux et al PNAS 2007, Ben Shem PNAS 2007, Vinothkumar et al EMBO J 2010, Vinothkumar and Freman 2013) having little in common with a protein substrate. In this manuscript, we present a state-of-the-art molecular dynamics (MD) model based on crystal structure with a peptide inhibitor demonstrably binding in a substrate-like mode. We believe the model is valid, because it was generated independently from, yet it is consistent with a number of published experimental results (beta-lactam complex structures and the cross-linking of TMDs 2 and 5 in a W236C/F153C mutant by the Ha group, position of the P1 carbonyl oxygen and the DFP complex structure). Figure S8 also provides graphs and tables that, in addition, support the validity of our molecular model. This is probably the most rigorous approach that has been taken so far in the field. Our MD model of the P4-P3' segment of a substrate exploits and adds value to our crystallographic data, and we thus believe that is an important and useful element of the paper.

The respective passage (p. 10) was reformulated for clarity as follows:

„Throughout MD simulations, H-bonds between the L3/L5 loop and the substrate backbone, as present in the crystal structure (Fig. 3B), were retained (Fig. S8B). Furthermore, we observed (i) the formation of H-bonds between the catalytic dyad residues (ii) the scissile bond carbonyl carbon and the S201 side chain oxygen coming into close spatial proximity compatible with nucleophilic attack, and (iii) formation of H-bonds between the P1 carbonyl oxygen and residues thought to form the oxyanion hole (Fig. S8B). The interactions (iii) involved mainly the H-bond by the N154 side chain nitrogen as the most stable one and by the S201 main chain amide. The latter H-bond was transient, while the previously observed H-bond to L200 main chain amide (Vinothkumar et al, 2010) could not be detected. During MD simulations, His150 transiently flipped back into the position it adopts in the unliganded enzyme (data not shown), suggesting that H150 (and maybe also L200) may hydrogen-bond to the negatively charged oxyanion that forms in the tetrahedral intermediate (but is absent from the Michaelis complex). Overall, the carbonyl oxygen of the P1 residue adopts a similar orientation in our MD simulations as found in the complex structure with diisopropylfluorophosphonate (DFP), deemed to mimic the tetrahedral intermediate (Xue & Ha, 2012) (Fig. S8C). This finding makes us confident that our MD model of the Michaelis complex (Fig. 6A) is realistic, allowing us to examine the interactions of the prime side residues with GlpG and estimate the likely exit position of the unwound C-terminus of the substrate from the body of GlpG.“

Original comment:

(8) Page 12-13: The homology modeling of TatA TMD substrate and docking to GlpG is highly speculative and should be excluded. Why is this added only to the discussion section?

Response:

We emphasise that this is a speculative structure-based conceptual model, and precisely for this reason has it been placed in the Discussion section. We believe that it is a reasonable extension of the interpretation of our structural data, and we have ensured that its speculative nature has been clearly stated.(p. 10):„ To propose a structure-based conceptual model of a full transmembrane substrate complex with GlpG we took advantage of the recent solution NMR structure of *E. coli*

TatA (Rodriguez et al, 2013). A homology model of *P. stuartii* TatA that we generated shows that the region spanning residues P13 (P4' position) to F27 is α -helical and about 22 Å long (Rodriguez et al, 2013). The estimated hydrophobic thickness of GlpG molecule from the point of exit of the P3' residue to the cytoplasmic boundary of the membrane is about 13 Å (Fig. 6B), and manual docking of *P. stuartii* TatA TMD region P13 (P4' position) to F27 into a representative structure of the Michaelis complex model suggests that the TatA TMD would 'stick out' of the membrane. ..."

Original comment:

(9) Page 14: It is stated "Although the whole cavity forming the S1 subsite is quite large, only small residues, such as alanine, cysteine or serine are accepted in the P1 position of the substrate. One explanation for this could be that larger side chains could interfere with the stability of the water molecules in the water retention site and thus inhibit the deacylation step." But the CMK inhibitors with large P1 residues are also not accepted and these do not undergo deacylation in working as inhibitors. This seems to be an inconsistency.

Response:

It is absolutely correct that the inhibitors do not require deacylation to work. This issue is more complex than we managed to capture in the space-restricted conditions of the initial submission. We now explain it in more depth. First of all, to visualise the properties of the S1 subsite and water retention site cavity we have generated new Fig. S6 in Supplemental Results, which shows the electrostatic surface potential of the cavity.

We first say on p. 8 of Results:

"The S1 subsite is the proximal part of a deeper cavity, whose distal part has a strongly hydrophilic character with negative surface electrostatic potential (Fig. S6) and contains three conspicuous conserved water molecules present in all structures of GlpG from different crystallization conditions and space groups (Ben-Shem et al, 2007; Vinothkumar, 2011; Wang et al, 2006)."

And we continue a detailed explanation of the S1 subsite/P1 residue issue in a rewritten paragraph of the Discussion. The bottom line is that there may be various reasons for why particular amino acids are not accepted into the S1 subsite. This could be because of steric reasons (F or W simply will not fit into the S1 subsite cavity, electrostatic reasons (residues with a negative electrostatic potential or hydrophobic residues will be repelled by the negative electrostatic potential of the cavity surface), or other options detailed below in the revised text:

"First, the 'water retention site' forms a continuous cavity with the S1 subsite of GlpG. Although the whole cavity is quite large, only alanine and to a much lower extent also cysteine or serine are accepted in the P1 position of the substrate. One explanation could be that the strongly negative electrostatic potential of this cavity (Fig. S6) disfavours binding of negatively charged residues and residues with longer aliphatic side chains than that of alanine. Polar natural amino acids are likely to be either too large to be accommodated (K, R, H) or might engage in hydrogen bonds to the water molecules inside the retention site, thus perturbing the described dynamic hydrogen bonding network (Zhou et al, 2012). Such interference could result in i) structural destabilization of the enzyme-substrate complex or ii) impaired catalysis as water molecules may not effectively access the catalytic site to be used in the deacylation step. The latter mechanism is experimentally testable, since one would predict that a substrate with a P1 residue of a suitable character larger than an alanine could be trapped at the acyl-enzyme stage, bound to the catalytic serine. However, given the structural restraints of the cavity and the structural properties of genetically encoded amino acids, testing this hypothesis might require the use of unnatural amino acids. Our structural analyses also rationalise why glycine is poorly tolerated in the P1 position of a substrate and the corresponding peptidyl-CMK. The poor tolerance cannot be due to steric hindrance, because glycine has no side chain, but it can be caused by a higher degree of rotational freedom endowed by glycine, which could prevent optimal alignment of the ligand's polypeptide chain for hydrogen bonding to the L3 loop backbone in a parallel β strand and productive exposure of the scissile bond to the catalytic residues"

Referee #3:

This referee is very positive about our manuscript, asserting that „this is a well-executed study that provides several new insights...“ and lists only four minor points to be addressed.

Original comment:

*Page 7 - The meaning of the following phrase is not clear, and should be reworded:
"Its part more distal to the P1 side chain, although buried in the core of the protein,..."*

Response:

We have now reworded this passage to increase clarity:

„The S1 subsite is the proximal part of a deeper cavity, whose distal part has a strongly hydrophilic character with negative surface electrostatic potential (Fig. S6) and contains three conspicuous conserved water molecules present in all structures of GlpG from different crystallization conditions and space groups...“

Original comment:

Page 14 - It is suggested that the reason the S1 pocket is not able to accommodate a larger side chain is that it would interfere with the stability of water molecules in the water retention site, and therefore inhibit deacylation. However, this would suggest that an acyl-enzyme intermediate could be trapped, an outcome that should give rise to a new species in the gel-based assay use to screen this substrate library yet was not observed. This should be acknowledged, or this speculation removed.

Response:

We have not observed any new species arising in our radioactive assay, and, given the assay format, we might not have observed them for two reasons, even if they had formed. First, if deacylation is inhibited but not completely blocked, the acyl-enzyme intermediate may be a minor component of the reaction mixture. Second, since substrate concentration in this radioactive assay is relatively low, less than about 1 μ M (Strisovsky, unpublished), and the peptide that would have been trapped in the acyl enzyme intermediate would have only one radioactive Met as opposed to the other cleavage product that harbours 4 labelled Met residues. The signal from the acylenzyme may thus be too weak to be detected.

As explained in our response to the last comment by referee #2, there are probably several structural and physico-chemical reasons why only small residues such as Ala, Ser or Cys are preferred at the P1 position. Specific detection of a ‘trapped’ acyl-enzyme intermediate might thus require the use of unnatural amino acids such as homoserine, 2,5-diaminopentanoic or 2,4-diaminobutyric acid in the P1 position of the substrate/inhibitor and mass-spectrometry, which is probably beyond the scope of this manuscript.

Original comment:

Figure 4A - Error bars are implied in the figure legend, but not shown in the figure.

Response:

This was a typo, which has been now corrected by correcting Fig. 4A legend to:

“The assays have been conducted three times independently and representative data are shown (source data in Fig. S7).”

Original comment:

Supplemental Figure S7 - Panel C appears to have been omitted, although it is described in the text and figure legend.

Response:

This omission was unfortunate and has now been corrected. This figure has now become Fig. S8C due to the insertion of a new supplementary figure S6.

Accept letter

19 August 2014

Thank you for submitting your final revised manuscript for our consideration. We have now been able to assess it as well as your responses to the original comments, and I am pleased to inform you that we have now accepted it for publication in *The EMBO Journal*.