bioRxiv preprint doi: https://doi.org/10.1101/324517; this version posted May 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Cis-nonPro Peptides: Genuine Occurrences and their Functional Roles

Christopher J. Williams, Lizbeth L. Videau, Bradley J. Hintze,
David C. Richardson, Jane S. Richardson

Department of Biochemistry, Duke University Medical Center, Durham NC 27710
Contact email: dcrjsr@kinemage.biochem.duke.edu

Abstract

Cis-nonPro peptides, a very rare feature in protein structures, are of considerable importance for two opposite reasons. On one hand, their genuine occurrences are mostly found at sites critical to biological function, from the active sites of carbohydrate enzymes to rare adjacent-residue disulfide bonds. On the other hand, a *cis*-nonPro can easily be misfit into weak or ambiguous electron density, which has led to a high incidence of unjustified *cis*-nonPro over the last decade. This paper uses the greatly expanded crystallographic data and newly stringent quality-filtering to identify the genuine occurrences and survey both individual examples and broad patterns of their functionality. The accompanying paper describes the problem of *cis*-nonPro over-use, including its causes, validation, and correction.

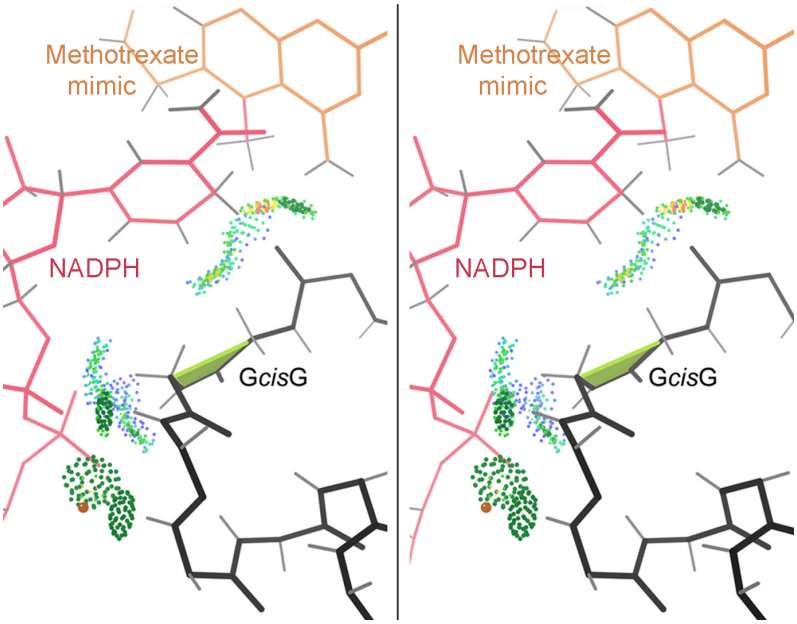
We explain the procedure developed to identify genuine *cis*-nonPro examples with almost no false positives, including the new observation that peptides with a glycine on one side or the other need extra care to avoid mis-assignment as *cis*-nonPro. We then survey a sample of the varied functional roles and structural contexts of *cis*-nonPro, emphasizing aspects not previously covered systematically: the preferred occurrence at β-strand ends in TIM barrel structures, the concentration of occurrence in proteins that process, bind, or contain carbohydrates, and the resulting complications in defining a simple occurrence frequency.

Introduction

Peptide bonds in protein structures have a partial double-bond character, which keeps them close to planar. The overwhelming majority adopt the *trans* backbone conformation, with the ω dihedral angle near 180°, while some are *cis* with ω near 0°. Peptides preceding prolines have somewhat less unfavorable energy difference and a lower barrier to *cis/trans* rotation (reviewed in Pal 1999), and the reported occurrence frequency of *cis*-Pro has converged over the years to a bit over 5% (Stewart 1990; MacArthur 1991; Jabs 1999), more common in β-sheet proteins than in helical ones. *Cis*-Pro play structural roles, as exposed turns or to accommodate tightly-packed regions in the interior. They also play static functional roles such as positioning active-site residues in carbonic anhydrase and in pectate lyase (Videau 2004). Such roles give them rather high conservation within a protein family (Lorenzen 2005). On the dynamic side, a need to achieve this rarer isomer produces a slow step in the protein folding process (Brandts 1975; Schmid 1992), and there are well-studied proline isomerase enzymes which aid in that step (Wawra 2006). The slow *cis/trans* conversion has been exploited by evolution for its suitability in biological timing circuits (Lu 2007).

bioRxiv preprint doi: https://doi.org/10.1101/324517; this version posted May 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In contrast, genuine *cis* peptides preceding non-proline residues are extremely scarce. Historically, the first examples were three *cis*-nonPro near the active site of carboxypeptidase A (Rees 1981) and the Gly-*cis*-Gly peptide that helps bind both NADPH and inhibitors in dihydrofolate reductase(Bolin 1982) . Figure 1 shows that site at 1.09Å resolution in the later 1kms human DHFR (Klon 2002).



Results

Reliability filtering for genuine occurrences

Any search for an unusual conformation needs quality filtering, at the residue level as well as the file level, in order to achieve a robust structural-bioinformatics treatment. Even more rigorous filtering is required for *cis*-nonPro peptides, given the recent epidemic of over-use (Croll 2015) in poor density at all resolutions. Therefore, as well as our usual steric and geometrical criteria, for this analysis we have developed additional criteria based on electron density quality (see Methods section and accompanying paper), modified from those used in our recent sidechain rotamer analysis (Hintze 2016). This added filter will reject some genuine cases, but has been shown to accept extremely few incorrect cases, and so is effective for choosing a list of reliably correct *cis*-nonPro examples for evaluating factors that affect occurrence preferences and for studying the real functional roles of this conformation.

The resulting reliability-filtered list contains 447 *cis*-nonPro peptides in 388 different protein chains, compiled from the Top8000 reference dataset of 6765 protein chains at < 2.0Å resolution, < 50% sequence homology, and with deposited diffraction data. This *cis*-nonPro list, with annotations, is given in Table S1 of the supplementary material and is also on GitHub at Rlabduke. It forms the primary basis of the analyses in this paper, with a few exceptions that investigate relationships or conservation across a wider range.

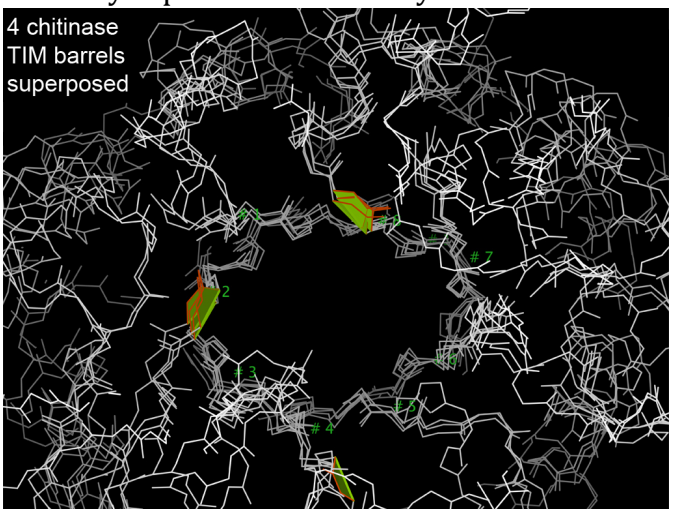
Preferential occurrence of *cis*-nonPro in carbohydrate-active proteins

It has been known for quite some time that a number of carbohydrate enzymes make use of *cis*-nonPro conformations (Jabs 1999), and we are now in a position to judge whether this apparent connection is actually significant. Our source for automated identification of carbohydrate-processing or binding proteins is inclusion on the CAZy web site (Carbohydrate Active enZymes; http://www.cazy.org; Lombard 2014). 99 of our 388 distinct protein chains with a genuine *cis*-nonPro (26%) are in CAZy, while only 6% of the 6765 total chains in our reference dataset (the Top8000_50%_SF) are in CAZy. Carbohydrate-active protein chains are thus over-represented by a factor of >4 in this list. Of the 447 reliably genuine *cis*-nonPro peptides, 144 examples (32%) occur in proteins that are on CAZy. Another way of expressing this contrast is that 76% (33/44) of the protein chains with more than one authenticated *cis*-nonPro peptide are in CAZy. Those 33 chains are only 0.49% of the reference dataset, but they

bioRxiv preprint doi: https://doi.org/10.1101/324517; this version posted May 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

contain 77 (17.2%) of the *cis*-nonPro peptides, an over-representation of 35-fold. Thus carbohydrate-active proteins are very much more likely than other proteins to have more than a single genuine *cis*-nonPro per chain, as well as much more likely to have any at all.

Of the 44 protein chains with > 1 authenticated *cis*-nonPro, 15 (42%) are chitinases or chitinase-like: 9 of them have three *cis*-nonPros and 6 have two (see Table S1). We found a later-deposited protein chain with 4 genuine *cis*-nonPro, also a chitinase: 3wd0. Chitinases are remarkably widespread and sequence-diverse throughout phyla, providing either nutritional use or protection against chitin-containing arthropod and especially fungal organisms. Their diversity explains how so many of them can occur in a dataset with < 50% sequence homology.



The positions of *cis*-nonPro peptides in a sample of chitinases are shown in Figure 2. All chitinases have a Trp-X *cis*-nonPro at the C-terminal end of β strand 8 in their TIM-barrel fold, where the Trp is well packed and seems to act by positioning the backbone and residue X, which is usually the catalytic acid. In the triple *cis*-nonPro chitinases, the other two well-conserved *cis*-nonPros are at or near the ends of β strands 2 and 4, and are involved in substrate binding (Terwissa van Scheltinga 1996).

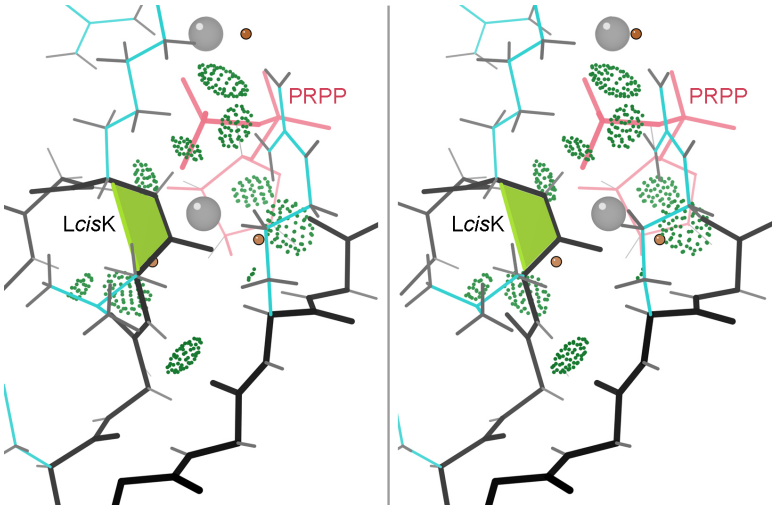
Besides the chitinases, chains with three *cis*-nonPro peptides include two β-galactosidases (1yq2 and 3cmg), in CAZy, and three carboxypeptidases (2piy, 3d4u, and 3i1u), not in CAZy but themselves modified with covalently-bound carbohydrate. The chains with two *cis*-nonPro peptides include a wider variety, 14 in and 8 out of CAZy.

The *cis*-nonPro-containing carbohydrate enzymes are unevenly distributed across the many structure/function families defined in CAZy. There are 17 *cis*-nonPro protein chains in family GH18 (glycoside hydrolase 18): the 15 chitinases and 2 xylanase inhibitors. There are 13 in family GH5, mostly mannanases, 11 in family GH1, mostly β-glucosidases, 8 in family GH10, all xylanases, 5 in family GH2, and 5 in different PL families (polysaccharide lyases). In contrast, the remaining 41 CAZy *cis*-nonPro chains are spread across 34 different families (see Table S1). The xylanases present an interesting combination of conservation versus divergence. The 8 in family GH18 all use a His-Thr *cis* peptide at the end of β-strand 3 of their TIM barrels. The 3 xylanases in other GH families each have a different *cis*-peptide sequence, located at the end of β-strand 1 (2ddx), β-strand 4 (2y8k), or β-strand 5 (1nof) of their TIM barrels.

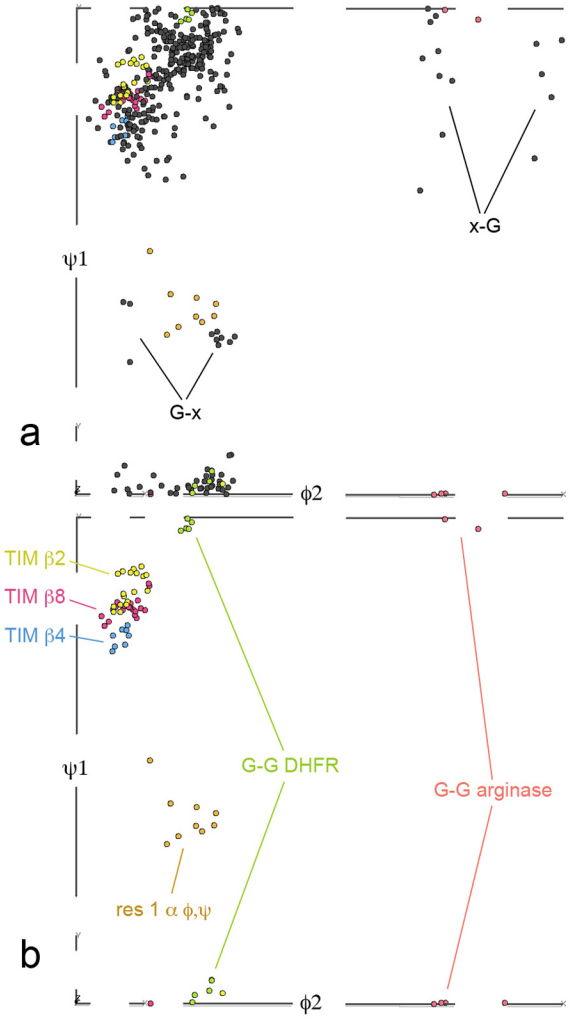
Lectins, or CBMs (Carbohydrate-Binding Modules), seem to be listed currently on CAZy only when they are part of a protein that also includes a carbohydrate enzyme, either as a separate domain or a separate chain. Therefore we did not have an automated way of identifying all of them uniformly. However, anecdotally, lectins such as the prototypical concanavalin A are fairly well represented on the *cis*-nonPro list (15 of them), but not as over-represented as the carbohydrate enzymes. The CBMs mostly have all-β antiparallel folds. The non-CAZy *cis*-nonPro proteins include 17 phosphoribosyl transferases, another highly sequence-diverse group (the PRT family) with only a 13-residue conserved sequence signature recognized. They share an essential *cis*-nonPro conserved in conformation but not in sequence,

bioRxiv preprint doi: https://doi.org/10.1101/324517; this version posted May 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

on the "PPi loop". As shown in Figure 3 for the 1.05Å 1fsg PRT structure, the backbone of the *cis*-nonPro and its adjacent peptides are used to bind a phosphate oxygen of PRPP and, through waters, a Mg ion (Sinha 2001). In at least some PRTs, a *trans* to *cis* change accompanies substrate binding (Shi 2002).



If the lectins, PRTs, and another 34 enzymes with some relationship to sugars are included, then 164 of the 388 *cis*-nonPro chains, or 42%, are carbohydrate related. Only three of these additional chains have two *cis*-nonPro, so 212 of the 477 *cis*-nonPro peptides, or 44%, are carbohydrate related. Since we made these additional assignments mainly from the titles and abstracts on the PDB site, they are not necessarily complete or correct, but even as an approximation this means that the predominance of carbohydrate relatedness is even stronger than the pure CAZy survey indicates.



<u>Preferential occurrence of *cis*-nonPro peptides in TIM-barrel proteins</u>

As well as their enzymatic activity on carbohydrates, another characteristic shared by many of the *cis*-nonPro containing protein chains is a $(\beta/\alpha)_8$ TIM barrel fold. Over half the relevant domains of the 99 CAZy proteins are TIM barrels, plus five $(\beta/\alpha)_7$ domains in phosphodiesterases. However, in the non-CAZy chains, TIM barrels are not especially over-represented, and indeed most TIM barrels (including the eponymous triose phosphate isomerase itself) do not include any *cis*-nonPro peptides. It seems, then, that the strong preference for $(\beta/\alpha)_8$ folds is probably not an independent factor, but is in some way correlated with presence in the carbohydrate-processing system. The more generic preference for β structure, and for location at or near the C-terminal end of a β strand, does seem to hold for non-CAZy as well as CAZy proteins.

<u>5-dimensional ϕ,ψ,ω,ϕ,ψ local conformational and sequence preferences</u>

A useful study tool is to plot datapoints for the genuine *cis*-nonPro examples in the 5-dimensional space of ϕ1, ψ1, ω, ϕ2, ψ2 selectable by their first and

bioRxiv preprint doi: https://doi.org/10.1101/324517; this version posted May 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

second amino-acid identities. These plots can be viewed 3 dimensions at a time in the Mage interactive graphics program (Richardson 2001). Not surprisingly, there is little spread from 0° in ω. ψ1, ϕ2 provides the most diagnostic projection, shown in Figure 4a for all *cis*-nonPro and in Figure 4b for selected clusters.

Overall, *cis*-nonPro peptides greatly prefer extended over helical conformations.

Trp-X *cis*-nonPro

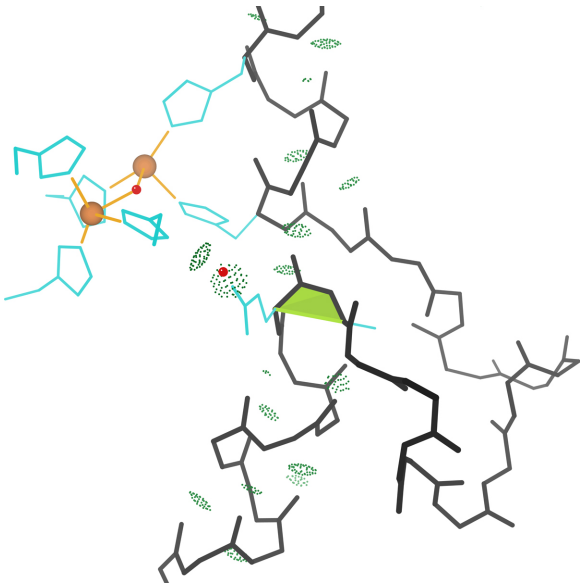
Trp and Cys are the least frequent amino acids in globular proteins (~1.3%). However, there are 49 Trp-X *cis*-nonPro in our dataset of 447, or 12%, an enrichment by a factor of 8, the highest ratio for any amino acid type. The preference is asymmetrical, with only 13 X-Trp and no Trp-Trp. This seems to be a functional rather than an energetic preference, since nearly all are on β strand 8 of a TIM-barrel carbohydrate-cleaving CAZy enzyme (14 chitinases, 7 glucosidases, 6 mannanases, 5 endoglucanase cellulases, and 6 other types in our dataset). Those ϕ,ψ values are all closely clustered in all 5 dimensions near the TIMβ8 points of Figure 4b, and the Trp sidechain adopts a **t-90** rotamer that lets it usually make some contact with the X sidechain (usually the active-site residue) and often with the adjacent β strand (see Figure 3b). Trp-Ser is the most common sequence, followed by Trp-Glu.

Five more Trp-X *cis*-nonPros are Trp-Thr in non-CAZy phosphodiesterases, at the end of β–strand 7 of 7-stranded TIM-like barrels. Only one Trp-X *cis*-nonPro is in a protein with no carbohydrate relationship at all: 1qgu nitrogenase.

First-residue helical ϕ,ψ

Helical ϕ,ψ values are very unusual for *cis*-nonPro, especially in the first amino-acid position. This loose cluster of 9 examples, however, turned out to have nothing else in common. One example (3frr A 46) has fairly good electron density but 5 geometry outliers, and is almost certainly wrong; it has been deleted from the overall list in Table S1. Most examples are some sort of turn, loop, or corner. Two are especially unusual and interesting. 2eab A His759-*cis*-Ala-*cis*-Pro761 is an unprecedented type of tight turn formed by a successive *cis*-nonPro and *cis*-Pro peptides, confirmed by definitive 1.12Å electron density. It has no evident functionality and we suspect it might not be conserved, but cannot tell because its crystal structure defined a new family (GH95) of fucosidases. (There is also a Gly-cis-Gly-cis-Pro224 in 3gne, but it is not an H-bonded turn.)

The one α-first case that has definite biological function is 3hhs chain A Glu-Ala354 (Figure 5) and chain B Glu-Ser352. They are conserved in related prophenoloxidase enzymes. The *cis*-nonPro forms the C-cap of one helix and makes two backbone H-bonds to the first turn of the next helix, separated by a short, meandering loop. The protruding *cis* peptide puts a bend of about 30° between the two helices. Glu 353 of the *cis*-nonPro stabilizes one of the 6 His ligands to the bi-nuclear Cu site, and the second helix contributes two more of those His ligands.

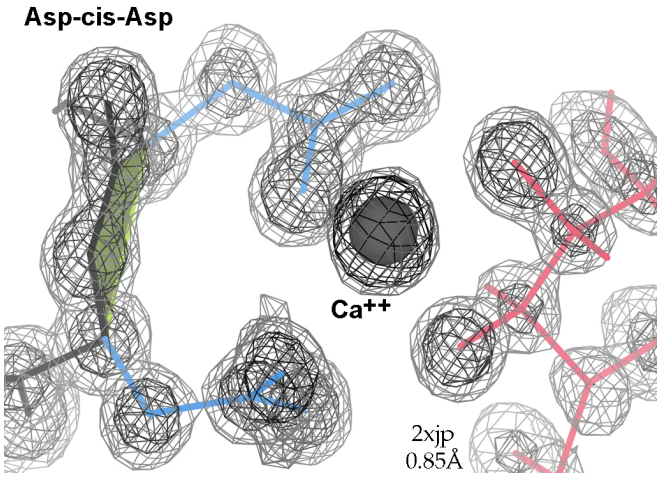


bioRxiv preprint doi: https://doi.org/10.1101/324517; this version posted May 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

DHFRs & other Gly-*cis*-Gly

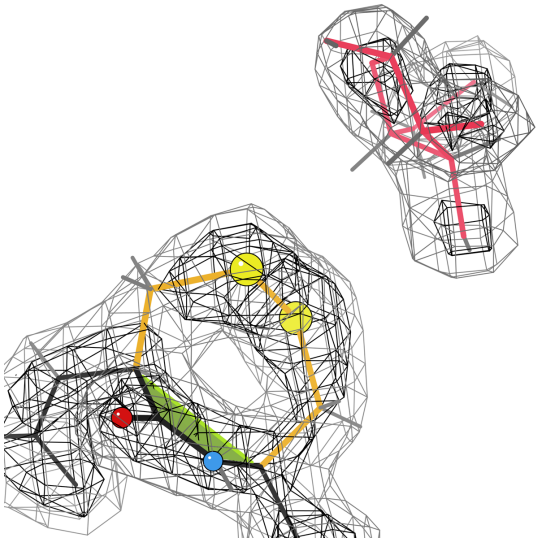
Gly is enriched generally by about 3-fold in *cis*-nonPro, at least partly because a Gly-containing peptide is much easier to fit incorrectly as *cis* and thus their numbers are inflated in unfiltered data. There is also a functional preference for Gly in some contexts, especially for Gly-Gly. Expectation for Gly-Gly would be 3.4 examples in our dataset, and there are actually 19. 10 of them are in dihydrofolate reductases (DHFRs) as a conserved and essential feature of the active site. Figure 1 illustrates the local conformation and interactions in the classic DHFR Gly-Gly *cis*-nonPro, showing how it is a major factor in binding both the NADPH cofactor and the methotrexate inhibitor, using H-bonds from the cis and surrounding peptides and van der Waals contacts from the essential Gly H "sidechains". These features are all conserved in DHFRs, as shown even most cleanly in the later structures of our dataset such as the 1.09Å 1kms human DHFR.

The other Gly-Gly *cis*-nonPros are quite varied, including 3 in arginase-superfamily enzymes and 3 in carbohydrate enzymes with β-helix folds. Their conformational clusters are in orange in Figure 4b, quite distinct from the DHFR cluster.



Asp-*cis*-Asp

Most of the nine Asp-Asp *cis*-nonPro peptides arrange their two sidechain carboxyl groups to form one side of a divalent-cation binding site, where the other side is bound by substrate. Figure 6 shows an example from 2xjp at 0.85Å resolution, with bound Ca^{++} and mannose. This flocculin enzyme in *Candida albicans* promotes the self-aggregation called flocculation, one of the steps in brewing beer. Four other Asp-*cis*-Asp examples bind the Zn^{++} (3ife, 3iib, 3pfe) or Mn^{++} (2pok) at the active site of a metallopeptidase. In the 2jdi F1-ATPase α-chain, just Asp269 binds the Mg^{++} through a water; in the β-chain, Asp 256 binds Mg^{++} similarly, but is an Asp-*cis*-Asn. 2rb7 is a *Desulfovibrio* metallopeptidase with no evident metal binding in the crystal and no publication, but the Asp-cis-Asp is part of a suggestive cluster of 3 Asp, 2 Glu, and 2 His deep in a cleft.



Cys-*cis*-Cys: vicinal disulfides

The *cis*-nonPro list includes two examples of Cys-*cis*-Cys, both of them SS-bonded as sequence-adjacent, or "vicinal", disulfides. Figure 7 shows the *cis* vicinal SS in 1wd4 arabinofuranosidase at 2.07Å resolution, where the disulfide makes van der Waals contact with bound arabinofuranose. The *trans* conformation is also possible, and actually more frequent, for vicinal disulfides. Therefore we have analyzed the occurrence patterns, the possible conformations (2 for *cis* and 2 for *trans*), and the

bioRxiv preprint doi: https://doi.org/10.1101/324517; this version posted May 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

varied functional roles of vicinal SS in a separate paper (Richardson 2017). They can bind ligands (usually the undecorated side of a ring, as in Fig. 7), stabilize structure, or provide the switch for a large conformational change.

Discussion

The most novel and striking conclusion of this study is also very puzzling. Why are *cis*-nonPro peptides selected by evolution about 5 times more often in carbohydrate-active enzymes than in other proteins, and multiple *cis*-nonPro >30 times as often? As far as we have seen, the *cis* peptide never itself performs the catalysis, and it seems to position a catalytic sidechain or to bind the carbohydrate by many quite distinct detailed strategies. Only a subset of CAZy enzymes, and extremely few other enzymes, take advantage of these *cis*-nonPro capabilities. Our best guess is that there are non-proline peptide isomerases that have been coupled into the systems that produce and control carbohydrate-related enzymes. A great many prolyl-peptide isomerases are known, but apparently a generic peptide *cis-trans* isomerase activity has so far been described only for DnaK (Schiene-Fischer 2002), and not with a carbohydrate connection. A search for such enzymes might well be productive and informative.

To support their biological functionality, some *cis*-nonPros position both sidechains (see Figures 6,7), while some position just one sidechain (Figures 2b, 5a). Other examples use interactions with the NH and/or CO of the *cis* peptide itself (Figures 1, 3), and those interactions can be either direct or through a water. The two sidechains and the functional groups of the peptide are approximately on opposite sides of the motif. *Cis*-nonPro almost always occur in well-ordered regions of the structure, usually with one side open for business and the other side held by tight contacts. One exception is vicinal disulfides, which can occur and function on very mobile loops, since the disulfide makes a *cis-trans* transition extremely difficult.

As usually true for rare motifs, nearly all genuine *cis*-nonPro are functionally important. They are rare because they are energetically quite unfavorable relative to trans, so they are not conserved by evolution unless they are useful and needed. Of many dozen examples examined, only a very few had no evident connection with function, such as the 4th one in the all-β domain of chitinase 3wd0 and the Asp-*cis*-Asp in 3das. There are also a few not obviously functional cis-Pro peptides on TIM-barrel strand ends, perhaps recognized by an over-enthusiastic generic isomerase. A major motivation for avoiding or fixing incorrect *cis*-nonPro peptides (as described in the accompanying paper) is to improve the signal-to-noise ratio for recognizing the genuine ones and investigating their functional roles.

Methods

The starting dataset of *cis*-nonPro examples is from a version of the "Top8000" that includes only the best chain in each RCSB PDB 50% homology cluster, requiring deposited diffraction data, and satisfying other quality criteria. Importantly, the dataset is further quality-filtered at the residue level, aiming to cull out incorrect or unjustified cases and leave only genuine *cis*-nonPro examples. In addition to automated filtering, over 100 cases were examined in 3D with their electron density and validation markup, which identified only four incorrect examples. Three of the four contained a Gly, which increases the likelihood of a *trans*-

bioRxiv preprint doi: https://doi.org/10.1101/324517; this version posted May 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to-*cis* misfit. More details of protocols and criteria are given in the Methods section of the accompanying paper.

Examination of individual examples was done in the KiNG display and modeling program (Chen 2009), using $2mF_o$-DF_c electron density maps and sometimes difference-density maps, plus model-validation markup from MolProbity (Chen 2010; Williams 2017); in most cases the literature references were also consulted. All figures except Figure 4 were produced in KiNG. The Mage graphics program (Richardson 2001) was used for the 5-dimensional dihedral-angle analysis, because of its cluster-defining functionality and because it can support the 52 lists per subgroup needed to select first- and second-residue sequence and conformation combinations for study.

Occurrence frequencies for the amino acids in *cis*-nonPro peptides were normalized relative to expectation by comparing them with frequencies in the general protein sequence population, as given in the UniProt knowledge base at http://www.ebi.ac.uk/uniprot/TrEMBLstats, version as of November 2017. As described in the text, carbohydrate-active enzymes were identified by their inclusion in the CAZy database at http://www.cazy.org, accessed over mid to late 2016.

Acknowledgements

This work was supported by National Institutes of Health grants P01-GM064210 Project IV to JSR and R01-GM073919 to DCR.

References

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr* **D66**:213-221

Bolin JT, Filman DJ, Matthews DA, Hamlin RC, Kraut J (1982) Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution: I. General features and binding of methotrexate, *J Biol Chem* **257**: 13650-13662

Brandts JF, Halvorsen HR, Brennan M (1975) Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues, *Biochemistry* **14**: 4953-4963

Chen VB, Davis IW, Richardson DC (2009). "KiNG (Kinemage, Next Generation): A versatile interactive molecular and scientific visualization program", *Protein Sci* **18**:2403-2409

Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr* **D66**: 12-21

Croll TI (2015) The rate of *cis-trans* conformation errors is increasing in low-resolution crystal structures, *Acta Crystallogr* **D71**: 706-709

Hintze BJ, Lewis SM, Richardson JS, Richardson DC (2016) MolProbity's ultimate rotamer-library distributions for model validation, *Proteins: Struct Func Bioinf* **84**, online since Mar

bioRxiv preprint doi: https://doi.org/10.1101/324517; this version posted May 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

28, doi: 10.1002/prot.25039

Jabs A, Weiss MS, Hilgenfeld R (1999) Non-proline *cis* peptide bonds in proteins, *J Molec Biol* **286**: 291–304

Klon AE, Heroux A, Ross LJ, Pathak V, Johnson CA, Piper JR, Borhani DW (2002) Atomic structures of human dihydrofolate reductase complexed with NADPH and two lipophilic antifolates at 1.09 and 1.05Å resolution, *J Molec Biol* **320**: 677–693

Lu KP, Finn G, Lee TH, Nicholson LK (2007) Prolyl *cis-trans* isomerization as a molecular timer, *Nature Chem Biol* **3**: 619-629

Rees DC, Lewis M, Honzatko RB, Lipscomb WN, Hardman KD (1981) Zinc environment and cis peptide bonds in carboxypeptidase A at 1.75Å resolution, *Proc Natl Acad Sci USA* **78**: 3408-3412

MacArthur MW, Thornton JM (1991) Influence of proline residues on protein conformation, *J Molec Biol* **20**: 397-412

Matsumoto T, Nonaka T, Hashimoto M, Watanabe T, Mitsui Y (1000) Three-dimensional structure of the catalytic domain of chitinase A1 from *Bacillus circulans* WL-12 at very high resolution, *Proc Jpn Acad* **B75**: 269-274

Pal D, Chakrabarti P (1999) *Cis* peptide bonds in proteins: residues involved, their conformations, interactions, and locations, *J Molec Biol* **294**: 271-288

Richardson DC, Richardson JS (2001) "MAGE, PROBE, and Kinemages", Chap. 25.2.8 In International Tables for Crystallography Vol. F (Eds. M.G. Rossmann and E. Arnold), pp. 727-730. Kluwer Academic Publishers, The Netherlands, Dordrecht.

Richardson JS, Videau LL, Williams CJ, Richardson DC (2017) Broad analysis of vicinal disulfides: Occurrences, conformations with *cis* or with *trans* peptides, and functional roles including sugar binding, *J Mol Biol* **429**: 1321-1335

Schiene-Fischer C, Habazetti J, Schmid FX, Fischer G (2002) The hsp70 chaperone DnaK is a secondary amide peptide bond *cis-trans* isomerase, *Nature Struct Biol* **9**: 419-424

Schmid FX (1992) Kinetics of unfolding and refolding of single-domain proteins, chap 5 in Protein Folding, Ed. Creighton TE, W.H. Freeman & Co, pp 197-241

Stewart DE, Sarkar A, Wampler JE (1990) Occurrence and role of *cis* peptide bonds in protein structure, *J Molec Biol* **214**: 253-260

Terwissa van Scheltinga AC, Hennig M, Dijkstra BW (1996) The 1.8Å structure of hevamine, a plant chitinase/lysozyme, and analysis of the conserved sequence and structure motifs of glycosyl hydrolase family 18, *J Mol Biol* **262**: 243-257

Videau LL, Arendall WB III, Richardson JS (2004) The *cis*-Pro touch-turn: a rare motif preferred at functional sites, *Proteins: Struct Funct Genet* **56**: 298-309

Wawra S, Fischer G (2006) Amide *cis-trans* isomerization in peptides and proteins, chap 9 in *Cis-trans* isomerization in biochemistry, Ed. Dugave C, Wiley-VCH, Weinheim, pp 167-194

Weiss MS, Metzner HJ, Hilgenfeld R (1998a) Two non-proline *cis* peptide bonds may be important for factor XIII function, *FEBS Lett* **423**: 291–296

Weiss MS, Jabs A, Hilgenfeld R (1998b) Peptide bonds revisited, *Nature Struct Biol* **5**: 676

bioRxiv preprint doi: https://doi.org/10.1101/324517; this version posted May 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure Captions:

Figure 1: Stereo view of the genuine and conserved Gly-*cis*-Gly peptide involved in cofactor and inhibitor binding in dihydrofolate reductase, shown for the 1.09Å DHFR 1kms.

Figure 2: Superposition of four chitinase TIM barrels showing conservation of non-proline *cis*-peptides at the ends of beta strands 2, 4, and 8.

Figure 3: Stereo view of a genuine Leu-*cis*-Lys peptide with bound substrate and ions from the phosphoribosyl transferase structure 1fsg. *cis*-nonPro is conformationally conserved in PRT proteins.

Figure 4: Projections into the diagnostic ψ1, φ2 parameter space of high-dimensional *cis*-nonPro clusters for a) all *cis*-nonPro residues, and b) *cis*-nonPro cases discussed in this paper.

Figure 5: *Cis*-nonPro peptides only seldom occur at the first residue of a helix. In this case from prophenoloxidase 3hhs, the conformation stabilizes His residues that coordinate Cu ions.

Figure 6: Typical Asp-*cis*-Asp conformation forming a divalent cation binding site, from flocculin enzyme 2xjp.

Figure 7: A vicinal disulfide bond around a *cis* peptide from arabinofuranosidase 1wd4. The *cis* peptide configuration is possible, but not obligate, for vicinal disulfides. Here the disulfide binds a ring-shaped ligand.