

The *cis*-Pro Touch-Turn: A Rare Motif Preferred at Functional Sites

Lizabeth L. Videau, W. Bryan Arendall III, and Jane S. Richardson*

Department of Biochemistry, Duke University, Durham, North Carolina

ABSTRACT A new motif of three-dimensional (3D) protein structure is described, called the *cis*-Pro touch-turn. In this four-residue, three-peptide motif, the central peptide is *cis*. Residue 2, which precedes the proline, has ϕ , ψ values either in the “prePro region” of the Ramachandran plot near -130° , 75° or in the $L\alpha$ region near $+60^\circ$, $+60^\circ$. The $C\alpha(1)$ – $C\alpha(4)$ distance is 4–5 Å and the two flanking peptides lie parallel to one another, making van der Waals contact rather than a hydrogen bond. Apparently, this arrangement is locally unfavorable and therefore rare, usually occurring only if needed for biological function. Of the 12 examples in a 500-protein database, *cis*-Pro touch-turns are found at the catalytic sites of pectate lyase, Ni-Fe hydrogenase, glucoamylase, xylanase, and opine dehydrogenase and at the primary binding sites of ribonuclease H, type I DNA polymerase, ribotoxin, and phage gene 3 protein. In each of these protein families, the touch-turns serve different roles; their functional importance is supported by conservation and mutagenesis data. In analyzing the conservation patterns of these 3D motifs, new methods for in-depth quality evaluation of the structural bioinformatic data are employed to distinguish between significant exceptions and errors. *Proteins* 2004;56:298–309. © 2004 Wiley-Liss, Inc.

Key words: protein structure motif; structural bioinformatics; pectate lyase; Ni-Fe hydrogenase; opine dehydrogenase; glucoamylase; xylanase; ribonuclease H; DNA polymerase; ribotoxin; phage gene 3 protein

INTRODUCTION

Understanding of protein structure has been enhanced over the years by the characterization of common local folding motifs such as tight turns,¹ β bulges² and β hairpins,³ right-handed $\beta\alpha\beta$ or crossover connections,^{4,5} E-F hands,⁶ leucine zippers,⁷ helix N-caps and C-caps,^{8–10} etc. In combination with basic secondary structures and with larger scale folding patterns, they have proven very useful in describing, comparing, predicting, and designing proteins. However, those discoveries were made some time ago, and it seems probable that nearly all commonly occurring structural motifs have been recognized by now.

On the other hand, many cases have been described of specific sequence-structure motifs that each act as a

signature for a particular function such as a specific catalytic mechanism or the binding of a specific ligand; these are collected on Prosite (<http://us.expasy.org/prosite>). They can be powerful enough for recognition of quite distant examples. By definition, these functional sequence motifs apply to a fairly narrowly defined function and explicitly do not define a generic structure motif.

This study reports an idiomorphic new motif called the *cis*-Pro touch-turn, which we believe to be the first example of a new class of relatively rare, rather than common, local structural motifs. The *cis*-Pro touch-turn is rare, presumably because the local structure is unfavorable; it is seldom selected by evolution, except where its unusual properties are needed. Thus, it is a general rather than a specific indicator of catalytic or binding sites. The proline is in position 3 of the four-residue motif, with the central peptide *cis* rather than *trans* and the two flanking peptides in van der Waals contact with each other. Overall, between 6% and 10% of the prolines in proteins are *cis*, depending on the structure class¹¹; because of the high barrier to *cis*–*trans* isomerization, they often produce slow steps in folding.¹² The roles of *cis* prolines seem to be primarily structural; they are generally not very common at active sites, but some recent examples show *cis*–*trans* isomerization acting as a dynamic switch between functionally distinct conformational states.¹³

The *cis*-Pro touch-turns are a small and well-defined subset of *cis* prolines with interestingly different behavior. They were discovered in our recent intensive reexamination of the Ramachandran plots for proline residues and for pre-proline residues¹⁴ (any non-glycine, non-proline residue immediately preceding a proline in sequence). The following results define the *cis*-Pro touch-turn motif in general, discuss the function and conservation of specific examples, and describe how outliers to the patterns of conservation within a protein family were assessed for validity.

Grant sponsor: NIH, National Institute of General Medical Sciences; Grant numbers: GM-15000 and GM-61302.

*Correspondence to: J.S. Richardson, Department of Biochemistry, Duke University, 211 Nanaline Duke Building, Durham, NC 27710-3711. E-mail: jsr@kinemage.biochem.duke.edu

Received 7 October 2003; Revised 16 December 2003; Accepted 17 December 2003

Published online 28 April 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20101

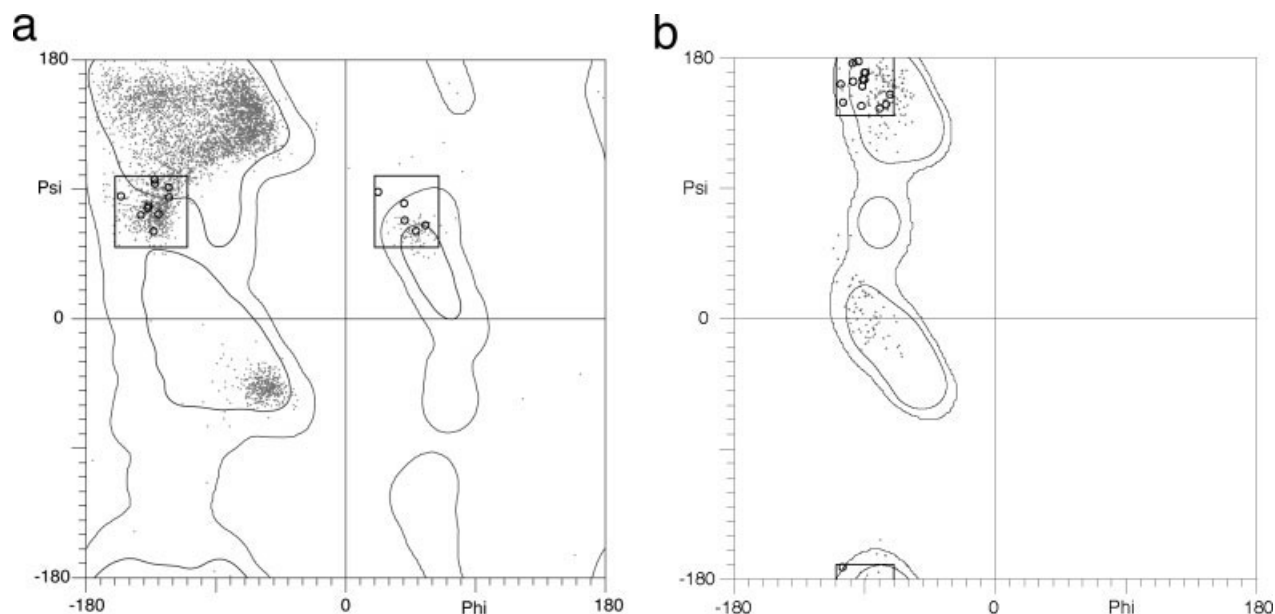


Fig. 1. Plots of ϕ , ψ for the two central residues in *cis*-Pro touch-turns. (a) All pre-proline (Pro-1) residues, both *trans* and *cis*, from (●) the Top500 data set versus contour outlines of the favored and allowed regions for general-case residues.¹⁴ (○) The 15 prototypical touch-turn pre-proline residues (Table I); the prePro and $L\alpha$ regions of ϕ , ψ space are within the boxed areas. (b) An analogous ϕ , ψ plot for the proline residues, with the region encompassing the touch-turn *cis*-Pro boxed in.

METHODS

The Top500 data set used here is taken from Lovell et al.¹⁴ It is a set of 500 nonredundant Protein Data Bank¹⁵ (PDB) structures at ≤ 1.8 Å resolution, quality filtered in a number of ways described in that paper. Ramachandran plots for the general, proline, and pre-proline (Pro-1) cases are also from Lovell et al.¹⁴ The tools used to check the reliability of a putative touch-turn in a specific structure are the Ramachandran outliers, side-chain rotamer criteria,¹⁶ C β deviation,¹⁴ and serious clashes in the all-atom contact display,¹⁷ as checked either individually or in the “MultiCrit” summary display by the MOLPROBITY service on our kinemage web site (<http://kinemage.biochem.duke.edu>).

A MySQL database¹⁸ of the Top500 was queried for residues preceding *cis* prolines and then those preceding *trans* prolines. Both sets of ϕ , ψ values were displayed similarly to Figure 1(a), but interactively, in MAGE.^{18,19} The pre-*cis*-Pro examples with ϕ , ψ in either the prePro region near -130° , 75° or in the $L\alpha$ region near $+60^\circ$, $+60^\circ$ were identified by clicking on them, from the residue number and PDB code on the resulting information line. A touch-turn is referred to here by its PDB code followed by a one-letter code for its four residues, with the proline residue number subscripted, for example, 1QCX RMP₂₃₈K.

For each example, a kinemage was created in PREKIN and displayed in MAGE to measure selected dihedrals, angles, and distances and to study both the local conformation and its environment. The four-residue *cis*-Pro touch-turns are defined by the central *cis* peptide, the short C α (1)–C α (4) distance, and the all-atom van der Waals contacts¹⁷ between flanking peptides (as calculated by Probe¹⁷). Those with pre-proline ϕ , ψ in the prePro region

have their flanking CO groups pointing in opposite directions, whereas those with pre-proline ϕ , ψ in the $L\alpha$ region have flanking CO groups approximately aligned. The Top500 was thus systematically searched for all *cis*-Pro touch-turns, either CO opposite or CO aligned. Several lower resolution examples (see Table I) were discovered in the comparison searches described below. When it became clear that many of the touch-turns were at active or binding sites, a literature search was done on each example.

The 21-residue sequence centered on each *cis*-Pro touch-turn was used to identify the Protein Information Resource (PIR) entry for that particular protein via an exact peptide match search in the PIR-Protein Sequence Database (PIR-PSD; see <http://pir.georgetown.edu>).²¹ A gapped-BLAST search²² then generated a list of PIR-PSD entries that were overall homologous “positive” to at least 35%. Conservation of the pertinent proline within the four-residue touch-turn sequence was determined from the subject to query sequence alignments. Each aligned PIR entry was checked for annotation of PDB structures covering the touch-turn portion; those were displayed in MAGE to evaluate the structural and functional context and to measure quantitative parameters for the putative touch-turn.

A search for more distantly related proteins was pursued via Dali²³ and FSSP.²⁴ The PDB ID for each Top500 *cis*-Pro touch-turn was submitted to the Dali “select structural neighbors” search. The ensuing list usually included structures in protein families not yet identified as containing touch-turns. A sample of these structures, taken from among the most divergent protein families and with greater than 10% identity indicated, were examined in MAGE for *cis* prolines and for the touch-turn configuration.

TABLE I. List of *cis*-Pro Touch-Turn Examples

	Touch-turn	Resolution	ϕ ψ :Xpr	Pro	Pro+1	$C\alpha_1-C\alpha_4$ Å	Virtual dihedral	Location	Site	Protein	Related PDB Files
Top500											
1QCX	RMP ₂₃₈ K	1.7	-142°, 72°	-106°, 161°	pP	4.5	-13°	β hx	A	Pectin/pectate lyase	1IDK, 1BN8, 1JTA, 1AIR, 1PCL, [1EE6]
1H2Rs	GCP ₁₅₁ P	1.4	-129°, 72°	-89°, 169°	pP	3.9	-9°	$\beta\alpha$	A	Ni-Fe hydrogenase	1E3D, 1CC1, 2FRV, 1FRF
1GAI	DNP ₄₆ D	1.7	41°, 80°	-75°, 147°	pP	4.5	41°	$\alpha\alpha$	—	Glucosylase	1AYX
1XNB	RSP ₇₅ L	1.49	42°, 68°	-98°, 163°	pP	4.2	17°	hp5~	A	Xylanase	1AXK, 1XYN, 1YNA, 1PVX, 1F5J, [1BK1]
1BG6	LNP ₁₀₉ G	1.8	59°, 65°	-90°, 169°	pP	4.1	21°	hp4	A	Opine dehydrogenase	1FOY
2RN2	GNP ₁₇ G	1.48	-122°, 91°	-105°, -172°	3_{10}	5.1	-8°	$\beta\alpha$	A	Ribonuclease H	1RIL
1XWL	TEP ₆₂₁ N	1.7	-136°, 78°	-90°, 164°	L β	4.1	3°	hp4	B	DNA polymerase I	2BDP, 2KFN, 1T7P, 1BGX, [1QSS]
1AQZ	KKP ₁₁₂ K	1.7	49°, 61°	-91°, 160°	pP	4.0	17°	$\beta\alpha$	B	Ribotoxin	1DE3, 1JBR
1G3P	TDP ₁₆₁ V	1.46	23°, 87°	-92°, 146°	pP	4.3	30°	lp	B	Phage gene 3 protein	2G3P
1BK0	LDP ₁₉₄ Y	1.3	-155°, 85°	-91°, 165°	pP	4.3	10°	hp5	B	Isopenicillin synthase	
3SIL	KAP ₁₃₆ D	1.05	-137°, 76°	-104°, 149°	pP	4.8	24°	lp	—	Sialidase, <i>Salmonella</i>	
Other											
1EE4	KKP ₂₄₈ Q	2.1	-122°, 84°	-98°, 176°	pP	5.0	-1°	lp	—	Importin/karyopherin α	1IAL
1TF4	GWP ₄₉₀ A	1.9	-132°, 93°	-79°, 145°	pP	4.5	19°	$\alpha\alpha$	B	Cellobiose E-4	2XAT, 1KRR
1KHR	GNP ₁₆₀ A	1.9	-133°, 60°	-94°, 177°	pP	4.1	-23°	lp	C	Galactoside acyltransferase	
									B, C		
Near touch turns											
3TDT	LEP ₁₇₀ L	2.0	-132°, 96°	-72°, 154°	pP	4.1	22°	hp4	B, C	THDP succinyltransferase	
1YTB	REP ₁₀₉ K	1.8	-157°, 117°	-66°, 155°	pP	4.7	24°	lp	A	TATA-binding protein	1CDW, 1VOK, 1PCZ
3GRS	VKP ₂₀₀ K	1.54	-142°, 114°	-74°, 145°	pP	4.6	15°	hp4	~B	Glutathione reductase	
1PLC	SHP ₃₇₅ P	1.54	-132°, 111°	-74°, 161°	pP	4.8	21°	hp4	~B	Plastocyanin; pseudoazurin	1IUZ, 1BQK, 1IBZ
2CUA	FVP ₁₆ S	1.33	-127°, 107°	-74°, 168°	pP	4.5	18°	hp5	~A	CuA domain of Cytochrome BA3	Distant: 1DLF, 1EPF
1MSP	YQP ₉₂ N	1.6	-124°, 105°	-77°, 164°	3_{10}	4.8	19°	sw	S	Major sperm protein	
	PNP ₉₄ I	1.6	-128°, 133°	-79°, 172°	β	6.1	31°	sw	S	Ribonuclease A	
			-145°, 111°	-85°, 161°	pP	5.7	25°	sw	S		
	TQP ₁₃ S	2.5	-147°, 117°	-58°, 151°	3_{10}	5.1	40°	sw	S		
	VDP ₅₇ P		-162°, 124°	-81°, 161°	3_{10}	5.6	24°	sw	S		
7RSA	GNP ₁₁₄ Y	1.26	-136°, 111°	-62°, 153°	pP	4.7	37°	hp4	—		

The ϕ , ψ for residue Pro+1 is described only by region, where pP stands for polyproline conformation; The virtual dihedral angle of the 4 C α atoms in the touch turn. Location is the structural location of the touch turn, where hp4 or hp5 has 4 or 5 residues between the last narrow H-bond pair of the β hairpin; lp is a longer loop between β strands; sw is an abrupt strand switch between the sheets of a β sandwich. Site is the functional role, with A for active site, B for binding site, C for domain or subunit contact, and S for a structural role; and ~ indicates either that the narrow pair of β H bonds is incompletely formed or that the touch turn is near the site but no functional involvement is evident.

The BLAST and Dali similarity searches led to identification of over 130 additional structures containing at least one *cis*-Pro touch-turn, including three new touch-turn families, plus examples of near-touch-turns.

The one issue that required a fairly subjective decision was setting the boundary between *cis*-Pro touch-turns and near-touch-turns. For most borders there is no problem: in the Ramachandran plot for pre-prolines, L α is an isolated cluster and the prePro region is a well-defined peninsula beneath the β region. The issue is choosing the ψ value that separates prototypical touch-turns with their pre-proline ϕ , ψ in the peninsula from near-touch-turns with pre-proline ϕ , ψ values in the β mainland. We settled on a dividing ψ of 100° after considering the Ramachandran plot shape, the degree of peptide contact, the range of ψ

values within a family, and the relative loss of functionality and conservation above 100°.

The SCOP²⁵ structural classification (<http://scop.mrc-lmb.cam.ac.uk/scop>) was used during and after the process of identifying the related group of proteins that contains instances of one individual *cis*-Pro touch-turn. The layout of the family, superfamily, and occasionally fold around the original example showed the extent of conserved occurrence and its boundaries. The nearest related SCOP examples just outside the occurrence region were always checked to provide negative controls. The SCOP layouts in this study were produced from a translated form of the SCOP electronic files (version 1.55) and entered into Excel for editing, annotation, and color coding. The pre-proline ϕ , ψ plot for 1QSS was produced in the KING viewer

online in MOLPROBITY. All other figures were produced in MAGE, with the images for Figure 3 rendered in Raster3D.²⁶

RESULTS

Table I summarizes the sequence, conformation, and other characteristics for 15 unrelated examples of *cis*-Pro touch-turns, plus another 10 near-touch-turn examples. Several particular details of the Ramachandran plot are important for describing this new structural feature. As shown by the data points in Figure 1(a), the ϕ , ψ distribution for pre-proline residues (those that precede a proline) differs greatly from the general non-glycine, non-proline ϕ , ψ distribution (outlined by thin, smoothed contours). The pre-proline backbone is more restricted by atomic clashes with the proline ring, allowing no population in the 3_{10} , $L3_{10}$, γ turn, or γ' regions. However, as first pointed out by Karplus,²⁷ an area around ϕ , ψ values of -130° , 75° , below the left edge of β , is preferentially populated by pre-proline residues. This “prePro region” in the upper left quadrant of the ϕ , ψ plot is boxed in Figure 1(a). Our much larger data set (see Methods section) confirms that pattern, having 70.6% pre-proline residues within the boxed prePro region. Because pre-proline residues make up only 4.5% of the total residues, this conformation is actually about 50 times more likely for a pre-proline than for a non-pre-proline residue. In contrast, glycine essentially never occurs in this ϕ , ψ region, whether it precedes proline or not.

If residues in the boxed prePro region in Figure 1(a) are divided into those that precede *trans* peptides (dots) and those that precede *cis* peptides (large open circles), then a very striking pattern is seen for the *cis* examples. The Top500 data set includes only seven cases of pre-*cis*-proline data points in this region of the Ramachandran plot, but each occurs within a remarkably consistent local structure motif not described before. As seen in Figure 2(a), the backbone changes direction by 180° around the *cis* peptide in a square corner that is reminiscent of an especially narrow tight turn, with $C\alpha_1$ and $C\alpha_4$ only 4–5 Å apart. However, instead of the flanking peptides pointing toward each other to allow a hydrogen bond, their planes lie parallel to one another at van der Waals contact distance (with the carbonyls pointing in opposite directions). Because of the peptide contact, we call this new motif the *cis*-Pro touch-turn; this version, with carbonyls opposite, is called a CO-opposite *cis*-Pro touch-turn. Figure 2(a) shows all seven examples superimposed, with all-atom contacts¹⁷ between the sides of the *cis*-Pro touch-turn for four of them. The upper portion of the contact is formed by the two flanking peptides, and contact is continued by the flanking side chains. Those side chains always touch each other or the backbone opposite, but their contacts do not all line up because the side-chain conformations can vary.

A second variant of the *cis*-Pro touch-turn also occurs, for which the conformation of the pre-proline residue is in the $L\alpha$ region (near $+60^\circ$, $+60^\circ$) rather than in the prePro region of the Ramachandran plot. These have a very similar shape [as seen in Fig. 2(b)] and the same sort of backbone contact as for the CO-opposite touch-turns, but

the two flanking carbonyls point in the same rather than opposite directions. There are five of these CO-aligned *cis*-Pro touch-turns in the Top500 data set, accounting for all of the pre-*cis*-proline residues in the $L\alpha$ region.

The *cis*-Pro touch-turn is defined by a ϕ , ψ pair (for the pre-proline residue) and the following ω near 0° , plus the backbone van der Waals contact. However, as seen in Figure 2, the region of overlap extends much farther than that. Figure 1(b) shows that the touch-turn proline ϕ , ψ is restricted to one corner of the region allowed for *cis*-proline (centered around $\phi = -90^\circ$, rather than the average $\phi = -70^\circ$ for general proline and -80° for *cis*-proline), whereas the next residue is always in the polyproline conformation (ϕ , ψ near -70° , $+150^\circ$) if it is not a glycine (see Table I). Nearly all examples are close to β strands along the sequence, but their structural environments vary widely: four are at the ends of fairly tight β hairpins, three are in longer loops between β strands, three are in β - α connections, one is in an otherwise regular parallel β helix, and one is between two α helices. In any of these contexts, the *cis* peptide and the parallel but touching peptides on either side produce unusual relative positioning of both side chains and backbone that cannot be reproduced by an ordinary tight turn or any other all-*trans* conformation.

The CO-opposite cases in Figure 2(a) with their pre-proline residue conformation in the prePro region of the Ramachandran plot are prototypical *cis*-Pro touch-turns with the flanking peptides optimally lined up for contact. Fairly similar near-touch-turn structures can occur with a normal β conformation ($\psi > 100^\circ$) for the pre-proline residue, but contact between the flanking peptides is smaller and more variable because those backbones are either sheared sideways relative to one another or opened wider than for a true touch-turn. A few such near-touch-turns are discussed below. The CO-aligned *cis*-Pro touch-turns in Figure 2(b) have optimal contact of their flanking peptides but an $L\alpha$ conformation of the pre-proline residue; because $+\phi$ conformations are more constrained, there is less variety. The only pre-*cis*-proline example found nearby but well outside the $L\alpha$ region (4XIS E186, at $+81^\circ$, $+102^\circ$) forms an H-bonded type VIa tight turn that is very different from a touch-turn. For both CO-aligned and CO-opposite touch-turns, the pre-proline ψ values are below 100° .

Sequence preferences are fairly weak for positions 1 and 4 of the *cis*-Pro touch-turns, whereas position 3 is always proline. The pre-proline residue (position 2) is 80% hydrophilic, 60% asparagine or aspartate or glutamine or glutamate, and never glycine, both in Table I and in related sequences.

Touch-Turn Examples and Their Functional Roles

Strikingly, it soon became clear that almost all of these *cis*-Pro touch-turns are located at functionally important sites, although there is no apparent commonality to the roles they play in these diverse proteins. *Cis*-Pro touch-turns are quite rare overall, and extremely so in general structure contexts. Judging by that rarity, they are presumably either energetically or kinetically much less favorable

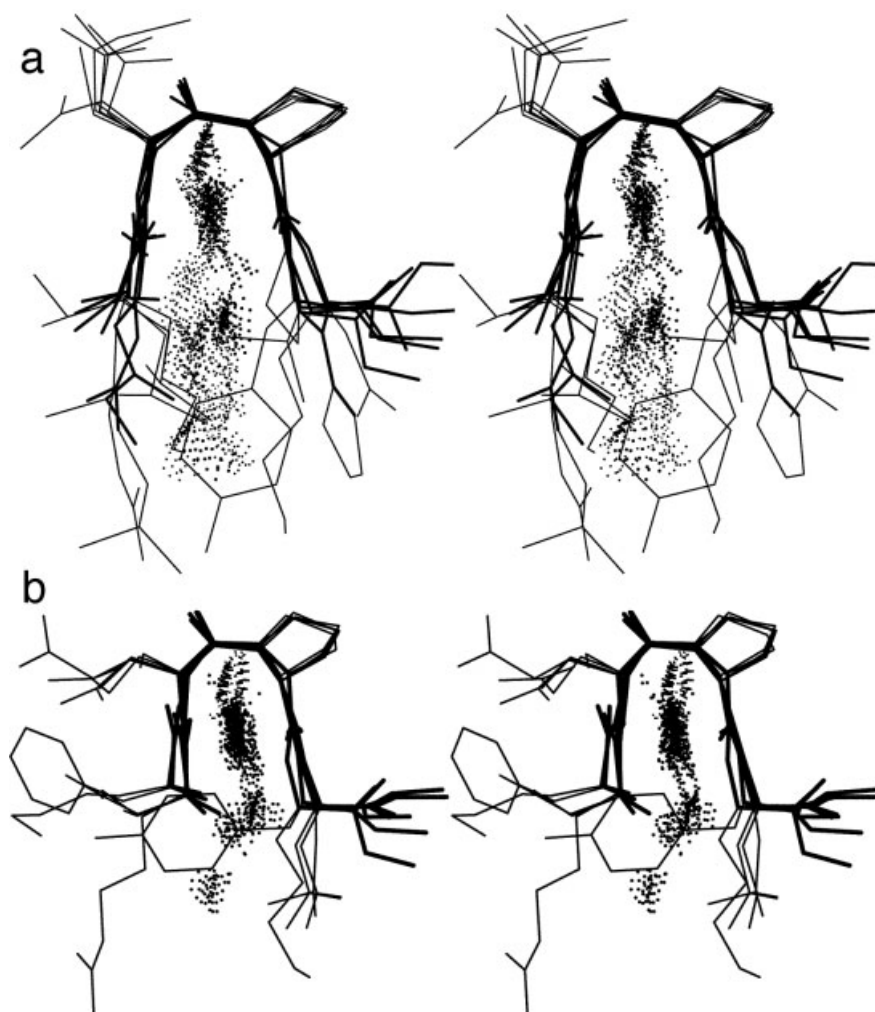


Fig. 2. Examples of *cis*-Pro touch-turns from the Top500 data set as superimposed in stereo. The backbone is shown as heavier lines, which are oriented with the central *cis* peptide across the top. (a) The seven CO-opposite touch-turns, with all-atom contact dots⁵⁶ between the sides for four of them. (b) The five CO-aligned touch-turns, with contacts for three of them.

than a normal tight turn, and less favorable even than other contexts for *cis* prolines. Therefore, it seems reasonable that they tend to occur only where there is a need for them. Thus, what makes this motif significant is not the unusual backbone conformation but is principally the notable fact that 9 of the 12 examples are found at active sites or binding sites, with functional importance confirmed by conservation or mutagenesis data. The 3 exceptions are in 3SIL, 1BK0,²⁸ and one of the two touch-turns in 1H2Rs.²⁹ All three are in surface loops fairly far from any recognized enzymatic or binding site, they show only limited if any sequence conservation, and their locations suggest no obvious functional role. The other 9 functionally relevant cases from the Top500 database will each be described: 5 at enzyme active sites and 4 at specific binding sites.

Enzymes 1

For the conserved active-site touch-turn in the pectin/pectate lyases (e.g., 1QCX RMP₂₃₈K), Arg236 is essential

for catalysis: even R236K is 10^3 lower in specific activity.^{30,31} Arg236 and Pro238 are conserved in all 43 related sequences found by BLAST, and the *cis*-Pro touch-turn conformation is the same in all six high-resolution structures of pectin lyase (A or B) and pectate lyase (A, C, or E[†]). Figure 3(a) shows the 1QCX touch-turn in the context of the surrounding turns of parallel β helices. The central Met 237 and Pro 238 chains lie within the hydrophobic core, and Arg236 reaches out from the long, concave side of the β helix. That concavity is formed by a line of residues (one on each strand), in α -helical conformation, which make all cross-strand hydrogen bonds but bend the β sheet. The *cis*-Pro touch-turn accentuates that bend, and the arginine has an extended side chain in hydrophobic contact with residue 239 (lysine or leucine). The flanking

[†]The touch-turn is absent in the 2.3 Å structure of a “high-alkaline” pectate lyase (1EE6) but the usual active site is also absent, and these differences are hard to evaluate without either a published article or deposited structure factors.

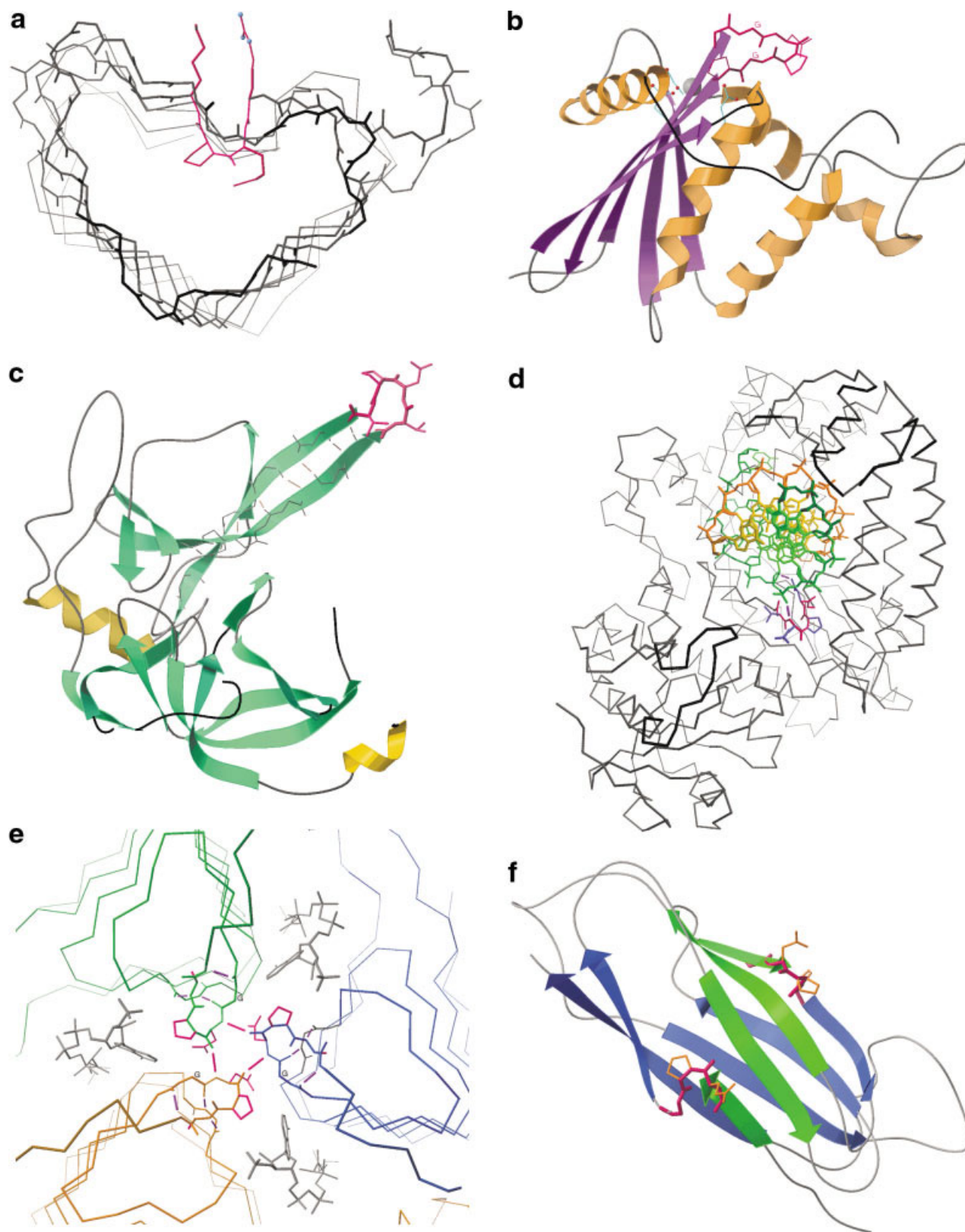


Fig. 3. Individual (pink) *cis*-Pro touch-turns at a variety of functional sites. (a) A slice through the β helix of pectin lyase (1QCX), where the touch-turn orients the catalytic Arg236. (b) Ribonuclease H (2RN2), where GNP₁₇G forms the loop binding the DNA/RNA hybrid just above the active site. (c) Phage gene 3 protein (1G3P) with the TDP₁₆₁V touch-turn poised at the end of the β hairpin that is probably responsible for pilus binding. (d) DNA polymerase complex (2BDP), with touch-turn H bonds to the template DNA strand. (e) Xenobiotic acyltransferase (2XAT),⁵⁷ where three touch-turns interact at the trimer interface, with (gray) the cofactors behind them. (f) The two sheet-switch touch-turns in major sperm protein (1MSP), forming Z-shaped strand connections on either side of the β sandwich.

peptides of the *cis*-Pro touch-turn make all four H bonds with the adjacent β strands up and down in the β helix. Although the touch-turn integrates fairly smoothly into the β helix, it must be relatively unfavorable, because the only other touch-turns in the eight families of parallel β helices (see below) occur at the ends of loops, not within the β helix. The conclusion, then, is that the *cis*-Pro touch-turn at the active site of pectin/pectate lyases is needed to position the essential arginine in its rather unusual arrangement.

Enzymes 2

In the Ni-Fe hydrogenases, a *cis*-Pro touch-turn (1H2Rs GCP₁₅₁P)²⁹ provides a cysteine ligand to the Fe₄S₄ cluster nearest the active site. The CP sequence is conserved in all 27 BLAST hits, and the *cis*-Pro touch-turn conformation is present in all four known structures of related enzymes, presumably needed to attain the correct ligand geometry in the specific context of this short $\beta\alpha$ loop.

Enzymes 3

In glucoamylase (1GAI),³² the DNP₄₆D touch-turn lies next to the bound substrate analog, with Tyr48 making extensive contact and residues 54–55 at the active site. Residues 41–43 preceding the touch-turn make three hydrogen bonds to an N-linked sugar. The PDY sequence and the touch-turn motif are conserved within the glucoamylase family, but not in the related cellulases. Although these enzymes are α - α barrels, the touch-turn is at the end of an irregular β hairpin.

Enzymes 4

In xylanase (1XNB), the RSP₇₅L *cis*-Pro touch-turn ends a β hairpin at the active site, where Glu78 is the catalytic nucleophile and Tyr69 and Tyr80 bind the substrate. The touch-turn is present in xylanase structures from 10 different species, but not in the related endoglucanases, which are classified in the same SCOP family but have considerably rearranged loops.

Enzymes 5

The LNP₁₀₉G touch-turn in 1BG6³³ is partially buried immediately beneath the norvaline dehydrogenase active site on a short $\beta\alpha$ loop that is reminiscent of the loop in 1H2Rs. The PG sequence is conserved in four of the seven related proteins found by BLAST. The *cis*-Pro touch-turn is apparently needed in one section of the opine dehydrogenases, but not for the entire family. In more distantly related dehydrogenases, one (1F0Y) has a near-touch-turn; it is in a similar but not equivalent position underneath the active site.

Binding 1

Three of the 12 touch-turn examples play critical roles at RNA or DNA binding sites, but each in a different way. As seen in Figure 3(b), the ribonuclease H *cis*-Pro touch-turn (2RN2 GNP₁₇G)³⁴ is exposed at the end of a bent antiparallel β ribbon with a conserved glycine-rich sequence known by mutagenesis and NMR to bind the RNA/DNA duplex.³⁴

The GNP GP sequence is conserved in the top 19 BLAST matches, and the *cis*-Pro touch-turn conformation is conserved in all known structures of type I RNase Hs. That region, however, is quite different in the structures of other types of RNase H. The flanking glycine residues allow the RNase H touch-turn to continue making side-by-side backbone van der Waals contact for four peptides rather than the usual two peptides; at that point the two chains bend abruptly by 90°, continuing down into a normal, H-bonded β ribbon in the main β sheet. The shape and nature of the RNase H substrate binding site would be very different without the unusual *cis*-Pro touch-turn arrangement.

Binding 2

In type I DNA polymerases, a touch-turn (1XWL TEP₆₂₁N)³⁵ helps bind the DNA template strand with both backbone and side-chain H bonds shown in purple for the polymerase/DNA complex in Figure 3(d). The proline and its following asparagine (the H-bonding side chain) are conserved in the type I DNA polymerase sequences. The touch-turn motif is conserved in the known type I and T7 phage structures, as discussed below, but that region is quite different in more distantly related polymerases.

Binding 3

The third nucleic acid binding example is 1AQZ³⁶ KKP₁₁₂K of ribotoxin and the related sarcin (1DE3), which are small toxins that cleave ribosomal RNA in the vulnerable sarcin/ricin loop. Substrate RNA binds on the convex, saucer-shaped face of ribotoxin, and the lysine-rich touch-turn loop sticks out at one side of that face. In the 1JBR³⁷ complex, all three touch-turn lysines interact directly or through water with the RNA loop at the characteristic kink in its double helix formed by the "S-motif."³⁸ The proline is conserved within the ribotoxin/sarcin family, and the touch-turn is also seen in the NMR structure of 1DE3. However, that loop has no lysine or proline in the distantly related ribonucleases with very different substrates, such as T1 or SA, or in the unrelated but functionally similar ricin.

Binding 4

Figure 3(c) shows a different sort of binding site example from domain N2 of the bacteriophage gene 3 protein (1G3P TDP₁₆₁V),³⁹ which initiates infection by binding to the tip of the F pilus and which is widely used for phage display experiments. The touch-turn forms the end of a long β ribbon that extends out from the structure. Sequence conservation is not helpful in this case, because the gene 3 proteins from filamentous phage are 99% identical and no other related proteins are found by BLAST or SCOP. There is some fold similarity with the gene 3 N1 domain and with PDZ domains, but they lack the β -hairpin extension and have quite different binding functions. Domain N2 is known to be responsible for pilus binding,⁴⁰ purportedly using the domain concavity and the threonine side chains on the N2 β ribbon,³⁹ one of which is in the touch-turn. We would suggest it is extremely likely that

the *cis*-Pro touch-turn is involved in that binding, because that unusual conformation is exposed at the very end of the β -hairpin loop where a normal tight turn would be more favorable if the only function were to form the turn.

Non-Top500 Database Examples

In the course of studying conservation of the *cis*-Pro touch-turns from the Top500, we happened to discover three additional examples at lower resolution, two at binding sites and two at domain or subunit contacts. The first is 1EE4 KKP₂₄₈Q from the mouse importin/karyopherin α , or “armadillo repeat.”⁴¹ This fold is an α - α toroid with the touch-turn at the end of an α - α loop that protrudes from the concave face of this crescent-shaped protein, contacting the nuclear localization sequence peptide bound in the crystal structure. The second is 1TF4 GWP₄₉₀A in the antiparallel β domain of cellobiose E-4,⁴² where the touch-turn forms a significant contact area of the domain-domain interface. The third non-Top500 example is GNP₁₆₀A in the 1KHR galactoside acyltransferase,⁴³ which forms both a trimer contact and a cofactor binding site. As shown in Figure 3(e) for the related 2XAT, three *cis*-Pro touch-turns interlock around the threefold subunit contact whereas their other sides help bind coenzyme A.

Thus, the above 12 distinct *cis*-Pro touch-turn examples are consistently found at either catalytic sites or important binding sites where their unusual three-dimensional (3D) structures could plausibly play functional roles, the importance of which is already confirmed in some cases by mutagenesis data. In all 12 examples, the enabling proline and functionally relevant residues are conserved, and the *cis*-Pro touch-turn structure is maintained throughout the entire protein family or a large subsection of the family. Outside of each related protein group sharing a functional touch-turn, that local region shows a quite different structure, presumably because an unfavorable conformation will not be retained once some other alternative is found to provide the functionality.

Near-Touch-Turns

As noted in the general description, there are also motifs similar to CO-opposite touch-turns but with the pre-proline residue ϕ , ψ in the β rather than the prePro Ramachandran region. These near-touch-turns have less peptide contact, longer $C\alpha_1$ - $C\alpha_4$ distances, and larger virtual dihedral angles on average. The boundary between touch-turns and near-touch-turns has been drawn at a pre-proline ψ value of 100° (see Methods section). In a sample of near-touch-turns, the only one at an active site is LEP₁₇₀L of the 3TDT THDP succinyltransferase.⁴⁴ It is a trimeric β -helix enzyme, but in this case the near-touch-turn is on a substrate binding loop at the interface between subunit pairs. 3GRS⁴⁵ SHP₃₇₅P is fairly near the glutathione reductase active site, but it is completely buried and poorly conserved. REP₁₀₉K and VKP₂₀₀K of the 1YTB TATA binding protein⁴⁶ are at equivalent edges of the DNA contact for the two domains and are conserved; the lysine interacts, but it is not clear why the near-touch-turn arrangement would be needed.

Some near-touch-turns appear to serve structural roles, such as providing specific interchain contacts (e.g., 3CHB KTP₉₃H in the cholera toxin pentamer)^{46b} or such as forming the kink that allows a β strand to switch between the two sheets of a β sandwich [i.e., the first half of the strand forms the edge of one β sheet, then it switches over to form an edge of the other sheet, as shown in Fig. 3(f)]. FVP₁₆S of 1PLC⁴⁷ and all the other plastocyanins, pseudoazurins, and nitrosocyanins use a *cis*-Pro in a near-touch-turn motif to make such a sheet switch, as do the CuA domain of cytochrome *ba3* (2CUA)⁴⁸ and both sandwich edges in the major sperm protein (1MSP).⁴⁹ All these proteins are Greek key β sandwiches, but from several different SCOP folds and superfamilies. The sheet switch ties together the two sheets midway, while a sheet switch is also a means for preventing β -sheet aggregation.⁵⁰ The rather similar azurin and immunoglobulin structures show, however, that sheet switches can also be made without *cis* peptides and even without prolines.

As an extreme but not uncommon case, Table I includes 7RSA⁵¹ GNP₁₁₄Y as an example of a near-touch-turn that is not very well conserved and does not appear to fill an important role. Thus, it seems that, in addition to having more open and less unusual conformations, the near-touch-turns are much less likely than prototypical *cis*-Pro touch-turns to be of functional significance.

Conformational Differences: Evolutionary or Dynamic Changes Versus Structural Errors

In assessing the conservation of a structural motif such as the *cis*-Pro touch-turn, an outlier that breaks the pattern is fairly often observed, which may simply be an error or may be an interesting and significant difference between environments or between proteins. For instance, Figure 4(a) shows the higher resolution examples of DNA polymerases, as organized by SCOP categories. The entries with *cis*-Pro touch-turns [unshaded areas, Fig. 4(a)] form a solid block covering almost all of the Pol 1 family, whereas all the structures outside that group [shaded areas, Fig. 4(a)] lack both the *cis*-Pro and the touch-turn motif. The boxed example (1QSS⁵²) in Figure 4(a), a Pol 1 from *Taq* with resolution equal to the highest for that species but with a *trans* proline, illustrates one of the central problems of such bioinformatic comparisons: is it a real exception to the pattern, or is it an error?

To decide that question, one can consult the validation tools provided by the MOLPROBITY web service¹⁴ to see whether that region of the structure is evaluated as sound or dubious. For 1QSS, the pre-proline Asp578 of the *trans* touch-turn mimic is listed as a Ramachandran outlier; on the resulting pre-proline ϕ , ψ plot [Fig. 4(b)], Asp578 lies in the lower right quadrant of the ϕ , ψ plot, far outside the 99.8% contour line for high quality pre-proline data points. MOLPROBITY's MultiCrit display summarizes on the 3D structure all outliers from the all-atom contact, Ramachandran, C β deviation,¹⁴ and side-chain rotamer criteria. In the zoomed-in stereo view of Figure 4(c), 1QSS 577-580 shows seven clusters of red spikes that mark bad steric clashes with an atomic overlap of ≥ 0.4 Å. There are no bad

a Conservation of cis P touch-turns in polymerases

PDB	res. (Å)	domain	SCOP class	fold	superfamily	family	protein	species	location	cis-Ptt
2kfn	2.03	A:519-928	Multi-domain	DNA/RNA pol	DNA/RNA pol	DNA pol I	DNA pol I (Klen)	E coli	Pro674	CO-aligned
1d8y	2.08	A:519-928	Multi-domain	DNA/RNA pol	DNA/RNA pol	DNA pol I	DNA pol I (Klen)	E coli	Pro674	CO-aligned
1krp	2.2	A:519-928	Multi-domain	DNA/RNA pol	DNA/RNA pol	DNA pol I	DNA pol I (Klen)	E coli	Pro674	CO-aligned
1bgx	2.3	T:423-832	Multi-domain	DNA/RNA pol	DNA/RNA pol	DNA pol I	DNA pol I (Klen)	T. aquaticus	Pro579	CO-aligned
11aq	2.4	423-832	Multi-domain	DNA/RNA pol	DNA/RNA pol	DNA pol I	DNA pol I (Klen)	T. aquaticus	Pro579	CO-aligned
1ktq	2.5	423-832	Multi-domain	DNA/RNA pol	DNA/RNA pol	DNA pol I	DNA pol I (Klen)	T. aquaticus	Pro579	CO-opposite
1qss	2.3	A:423-831	Multi-domain	DNA/RNA pol	DNA/RNA pol	DNA pol I	DNA pol I (Klen)	T. aquaticus	Pro579	trans-mimic
1xwl	1.7	469-876	Multi-domain	DNA/RNA pol	DNA/RNA pol	DNA pol I	DNA pol I (Klen)	B stear, strain	Pro621	CO-aligned
2bdp	1.8	A:469-876	Multi-domain	DNA/RNA pol	DNA/RNA pol	DNA pol I	DNA pol I (Klen)	B stear, strain	Pro621	CO-aligned
117p	2.2	A:211-704	Multi-domain	DNA/RNA pol	DNA/RNA pol	DNA pol I	T7 DNA pol	phage T7	Pro435	CO-aligned
1qht	2.1	A:348-750	Multi-domain	DNA/RNA pol	DNA/RNA pol	DNA pol I	T4-like DNA pol	A. Thermo_cc sp. 9or-7		no cisP
1d5a	2.4	A:348-756	Multi-domain	DNA/RNA pol	DNA/RNA pol	DNA pol I	T4-like DNA pol	A. Desulfurococcus tok		no cisP
1mmi	1.8	-	Multi-domain	DNA/RNA pol	DNA/RNA pol	Rev transcriptase	MMLV rev transcriptas	mm leuk virus		no cisPtt
1vrt	2.2	A:4-429	Multi-domain	DNA/RNA pol	DNA/RNA pol	Rev transcriptase	HIV rev transcriptase	HIV-1		no cisPtt
1oex	2.4	A:	Multi-domain	DNA/RNA pol	DNA/RNA pol	T7 RNA polymerase	T7 RNA pol	phage T7		no cisP
1c2p	1.9	A:	Multi-domain	DNA/RNA pol	DNA/RNA pol	Viral RNA-depender	RNA-dep RNAPol	HepatitisC virus		no cisP
1rpl	2.3	-	Multi-domain	Nucl transferases	Nucl transferase:	DNA pol beta-like	DNA pol beta	Rat		no cisP
115a	2.5	A:20-364	Multi-domain	Nucl transferases	Nucl transferase:	poly(A) pol	poly(A) pol	Cow		no cisPtt

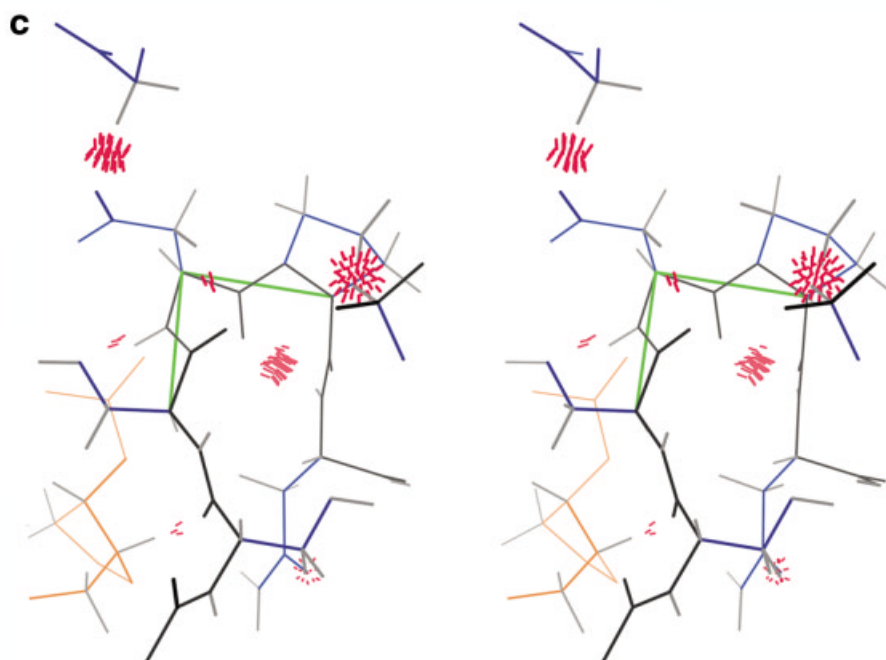
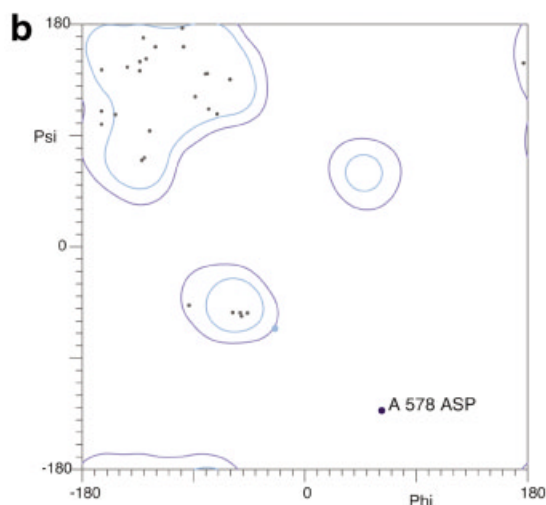


Fig. 4. A conservation analysis for the DNA polymerase touch-turn. (a) The SCOP classification for the related set of structures; those without *cis*-Pro touch-turns are shaded, and the trans touch-turn mimic (1QSS) is boxed in. (b) MolPROBITY pre-proline ϕ , ψ plot for 1QSS, flagging the pre-proline of the touch-turn mimic (Asp578) as a serious outlier. (c) MolPROBITY MultiCrit display in stereo for 1QSS zoomed in on the touch-turn mimic, with (red spikes) all-atom clashes,⁵⁶ (green lines) the Ramachandran outlier from (b), and the deviant τ angle at Asp578. The *trans*-Asp-Pro peptide is across the top in an orientation similar to Figure 2, and (gold) the DNA template strand lies behind.

rotamers or C β deviations in the touch-turn mimic, but the tetrahedral τ angle at the C α of Asp578 is distorted to 90°, which is >7 standard deviations from the ideal value of 111.2°.⁵³ In contrast to this *trans* touch-turn mimic, the CO-aligned *cis*-Pro touch-turn in 1XWL at higher resolution (Table I) has good ϕ , ψ and bond angle values and no serious clashes. Given the concentration of severe problems at the 1QSS touch-turn mimic, there is no doubt that the *trans* peptide represents a fitting error. Thus, we can conclude that the *cis*-Pro touch-turn is indeed completely conserved within this group of polymerases.

A second, more subtle feature for which one must assess reliability is the peptide flip that distinguishes CO-opposite from CO-aligned touch-turns, a difference with only very local effects. First of all, there are clearly real examples of both types, because each is seen in well-ordered regions of structures at 1.5 Å resolution and the touch-turn type is usually completely conserved within the related group. Of the three type switches seen between related but different proteins, EP₃₀ of 1H2Rs (1.4 Å) versus QP₃₀ of 1E3Da (1.8 Å) is almost certain to be real, giving an example of an evolutionary switch in peptide direction, although not for a touch-turn known to be functionally important.

Considering dynamic changes within one molecule, three type switches were seen among different structures for the same protein, two examples of which are probably incorrect. The 1EJY mouse importin α is at 2.9 Å resolution, where it is often difficult to reliably assign CO orientation. The DNA polymerase touch-turns from *Escherichia coli*, a bacillus, and T7 phage are all CO-aligned; and the 2BDP⁵⁴ structure in Figure 3(d) shows that the NH of the critical peptide hydrogen bonds with the DNA backbone. Two of the *T. aquaticus* structures at 2.5 Å (1KTQ and 5KTQ) are CO-opposite, but that orientation seems unconvincing because it prevents the H bond and instead puts the NH too close to another protein NH group. The third case, in the gene 3 protein, could well be real because the structures (1G3P, 2G3P) are at 1.9 Å or better and the *cis*-Pro touch-turns are at the end of a loop with no steric clashes or bad geometry. Such an example suggests at least the possibility of dynamic change for a touch-turn motif, which could be one of three sorts: a change between *trans* and *cis* similar to the examples reviewed by Andreotti¹³, or between CO-opposite and CO-aligned similar to the peptide flips seen for some tight turns,⁵⁵ or perhaps a relaxation from the close contacts of a *cis*-Pro touch-turn to the more open and skewed form of a near-touch-turn with β ϕ , ψ for the pre-proline residue.

DISCUSSION

The *cis*-Pro touch-turn constitutes a versatile new local structural motif that can help further our understanding and control of 3D protein structure. It is unusual in the fact that it potentially sheds new light on functional aspects of a dozen unrelated protein families, many of which are of major biological importance. We hope that new mutational tests will be conducted to confirm the predicted importance of the touch-turn and to discover the

specific role that makes conservation of this rather unfavorable structure worthwhile in each case. The range of peptide conformations and orientations for touch-turns and near-touch-turns raises the possibility that conformational transitions relating to functionality might occur, but so far no experimental data demonstrates such a case.

This study has necessitated the use of new analytical tools for reliability assessment, which we believe should be helpful in many other contexts of structural bioinformatics. Whenever comparisons or alignments are made, there are apt to be one or two exceptions that do not match what otherwise seems to be a very strong and convincing pattern. One should always try to assess the relative reliability of the data to decide whether such cases are interesting exceptions or are just errors. For 3D structural data, very easy evaluation tools are now available, as described in the Results section. Such analyses allowed us to cleanly show that *cis*-Pro touch-turns are entirely conserved within a protein family or one subsection of a family; then, when evolution finds another way to achieve that function, the touch-turn is completely lost.

In the most general sense, now that so many high-quality structures are available, we believe the *cis*-Pro touch-turn will be only the first in a series of local motifs that are rare but functionally important. The touch-turns could probably be recognized by NMR and in even fairly low-resolution X-ray structures, and their presence in a new structure would signal, with rather high probability, a functionally important region. These results imply that the significance of unusual structural idioms can actually come about because of, rather than in spite of, their rarity.

CONCLUSIONS

The *cis*-Pro touch-turn motif enables unusual positioning of both backbone and side-chain elements, about 75% of the examples making specific arrangements that are functionally important at enzyme active sites or at recognition and binding sites. Thus, not only can these motifs further illuminate structure/function relationships for a wide variety of well-studied protein families, but also their presence acts as a high probability locator of functional sites in new structures. This study shows the value and ease of using local data-quality indicators to distinguish interesting exceptions from probable errors within structural bioinformatic comparisons.

ACKNOWLEDGMENTS

We thank David Richardson for the kinemage graphics system and Ian Davis for the MOLPROBITY service. The motif analysis and the evaluation methods in this work were supported by two research grants from the National Institute of General Medical Sciences at the NIH.

REFERENCES

1. Venkatachalam CM. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers* 1968;6:1425–1436.
2. Richardson JS, Getzoff ED, Richardson DC. The beta bulge: a common small unit of non-repetitive protein structure. *Proc Natl Acad Sci USA* 1978;75:2574–2578.

3. Sibanda BL, Thornton JM. Beta-hairpin families in globular proteins. *Nature* 1985;316:170–174.
4. Sternberg MJE, Thornton JM. On the conformation of proteins: the handedness of the connection between parallel β -strands. *J Mol Biol* 1977;110:269–283.
5. Richardson JS. Handedness of crossover connections in beta sheets. *Proc Natl Acad Sci USA* 1976;73:2619–2623.
6. Kretsinger RH. Calcium-binding proteins. *Annu Rev Biochem* 1976;45:239–266.
7. Landschulz WH, Johnson PF, McKnight SL. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* 1988;240:1759–1764.
8. Richardson JS, Richardson DC. Amino acid preferences for specific locations at the ends of α -helices. *Science* 1988;240:1648–1652.
9. Presta LG, Rose GD. Helix signals in proteins. *Science* 1988;240:1632–1641.
10. Muñoz V, Blanco FJ, Serrano L. The hydrophobic-staple motif and a role for loop-residues in α -helix stability and protein folding. *Nat Struct Biol* 1995;2:380–385.
11. Richardson DC, Richardson JS. Principles and patterns of protein conformation. In: Fasman GD, editor. *Prediction of protein structure and the principles of protein conformation*. New York: Plenum; 1989. p 1–98.
12. Brandts JF, Halvorsen HR, Brennan M. Consideration of the possibility that the slow step in protein denaturation reactions is due to cis–trans isomerism of proline residues. *Biochemistry* 1975;14:4953–4963.
13. Andreotti AH. Native state proline isomerization—an intrinsic molecular switch. *Biochemistry* 2003;42:9515–9524.
14. Lovell SC, Davis IW, Arendall WB III, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by $C\alpha$ geometry: ϕ , ψ and $C\beta$ deviation. *Proteins* 2003;50:437–450.
15. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–242.
16. Lovell SC, Word JM, Richardson JS, Richardson DC. The penultimate rotamer library. *Proteins* 2000;40:389–408.
17. Word JM, Lovell SC, LaBean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS, Richardson DC. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogens. *J Mol Biol* 1999;285:1711–1733.
18. Widenius M, Axmark D. *MySQL Reference Manual*. New York: O'Reilly & Associates, Inc.; 2002. pp. 712.
19. Richardson DC, Richardson JS. The kinemage: a tool for scientific illustration. *Protein Sci* 1992;1:3–9.
20. Richardson JS, Richardson DC. *MAGE, PROBE, and kinemages*. In: Rossmann MG, Arnold E, editors. *International tables for crystallography*. Volume F: crystallography of biological macromolecules. Dordrecht: Kluwer Academic; 2002. p 727–730.
21. Wu CH, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu Z-Z, Ledley RS, Lewis KC, Mewes H-W, Orcutt BC, Suzek BE, Tsugita A, Vinayaka CR, Yeh L-SL, Zhang J, Barker WC. The Protein Information Resource: an integrated public resource of functional annotation of proteins. *Nucleic Acids Res* 2002;30:35–37.
22. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang W, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
23. Holm L, Sander C. DALI—A network tool for protein-structure comparison. *Trends in Biochem Sci* 1995;20:478–480.
24. Holm L, Sander C. Mapping the protein universe. *Science* 1996;273:595–602.
25. Murzin A, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 1995;247:536–540.
26. Merritt EA, Bacon DJ. Raster3D: photorealistic molecular graphics. *Methods Enzymol Macromol Crystallogr B* 1997;277:505–524.
27. Karplus PA. Experimentally observed conformation-dependent geometry and hidden strain in proteins. *Protein Sci* 1996;5:1406–1420.
28. Roach PL, Clifton IJ, Hensgens CM, Shibata N, Schofield CJ, Hajdu J, Baldwin JE. Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* 1997;387:827–830.
29. Higuchi Y, Ogata H, Miki K, Yasuoka N, Yagi T. Removal of the bridging ligand atom at the Ni-Fe active site of [Nife] hydrogenase upon reduction with H₂, as revealed by X-ray structure analysis at 1.4 Å resolution. *Structure* 1999;7:549–556.
30. Herron SR, Benen JA, Scavetta RD, Visser J, Jurnak F. Structure and function of pectic enzymes: virulence factors of plant pathogens. *Proc Natl Acad Sci USA* 2000;97:8762–8769.
31. Vitali J, Schick B, Kester HCM, Visser J, Jurnak F. The three-dimensional structure of *Aspergillus niger* pectin lyase B at 1.7-Å resolution. *Plant Physiol* 1998;116:69–80.
32. Aleshin AE, Stoffer B, Firsov LM, Svensson B, Honzatko RB. Crystallographic complexes of glucoamylase with maltooligosaccharide analogs: relationship of stereochemical distortions at the nonreducing end to the catalytic mechanism. *Biochemistry* 1996;35:8319–8328.
33. Britton KL, Asano Y, Rice DW. Crystal structure and active site location of *N*-(1-D carboxylethyl)-*L*-norvaline dehydrogenase. *Nature Struct Biol* 1998;5:593–601.
34. Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, Nakamura H, Ikehara M, Matsuzaki T, Morikawa K. Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J Mol Biol* 1992;223:1029–1052.
35. Kiefer JR, Mao C, Hansen CJ, Basehore SL, Hogrefe HH, Braman JC, Beese LS. Crystal structure of a thermostable bacillus DNA polymerase I large fragment at 2.1 Å resolution. *Structure* 1997;5:95–108.
36. Yang X, Moffat K. Insights into specificity of cleavage and mechanism of cell entry from the crystal structure of the highly specific aspergillus ribotoxin, restrictocin. *Structure* 1996;4:837.
37. Yang X, Gerczei T, Glover LT, Correll CC. Crystal structures of restrictocin–inhibitor complexes with implications for RNA recognition and base flipping. *Nature Struct Biol* 2001;8:968–973.
38. Leontis NB, Westhof E. A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. *J Mol Biol* 1998;283:571–583.
39. Lubkowski J, Henneke F, Plückthun A, Wlodawer A. The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. *Nature Struct Biol* 1998;5:140–147.
40. Deng L-W, Malik P, Perham RN. Interaction of the globular domains of pIII protein of filamentous bacteriophage fd with the F-pilus of *Escherichia coli*. *Virology* 1999;253:271–277.
41. Conti E, Kuriyan J. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. *Structure* 2000;8:329–338.
42. Sakon J, Irwin D, Wilson DB, Karplus PA. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Struct Biol* 1997;4:810–818.
43. Sugantino M, Roderick SL. Crystal structure of Vat(D): an acetyltransferase that inactivates streptogramin group A antibiotics. *Biochemistry* 2002;41:2209–2216.
44. Beaman TW, Blanchard JS, Roderick SL. The conformational change and active site structure of tetrahydrodipicolinate *N*-succinyltransferase. *Biochemistry* 1998;37:10363–10369.
45. Karplus PA, Schulz GE. Refined structure of glutathione reductase at 1.54 Å resolution. *J Mol Biol* 1987;195:701–729.
- 46a. Kim Y, Geiger JH, Hahn S, Sigler PB. Crystal structure of a yeast TBP/TATA-box complex. *Nature* 1993;365:512–520.
- 46b. Merritt EA, Kuhn P, Sarfaty S, Erbe JL, Holmes RK, Hol WJG. The 1.25 Å resolution refinement of the cholera toxin B-pentamer: evidence of peptide backbone strain at the receptor-binding site. *J Mol Biol* 1998;282:1043–1059.
47. Guss JM, Bartunik HD, Freeman HC. Accuracy and precision in protein structure analysis: restrained least-squares refinement of the structure of poplar plastocyanin at 1.33 Å resolution. *Acta Crystallogr B* 1992;48:790–811.
48. Williams PA, Blackburn NJ, Sanders D, Bellamy H, Stura EA, Fee JA, McRee DE. The CuA domain of *Thermus thermophilus* ba3-type cytochrome *c* oxidase at 1.6 Å resolution. *Nature Struct Biol* 1999;6:509–516.
49. Bullock TL, Roberts TM, Stewart M. 2.5 Å resolution crystal structure of the motile major sperm protein (MSP) of *Ascaris suum*. *J Mol Biol* 1996;263:284–296.
50. Richardson JS, Richardson DC. Natural β -sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci USA* 2002;99:2754–2759.
51. Wlodawer A, Svensson LA, Sjölin L, Gilliland GL. Structure of

- phosphate-free ribonuclease A refined at 1.26 Å. *Biochemistry* 1988;27:2705–2717.
52. Li Y, Mitaxov V, Waksman G. Structure-based design of novel Taq DNA polymerase with improved properties of dideoxynucleotide incorporation. *Proc Natl Acad Sci USA* 1999;96:9491.
 53. Engh RA, Huber R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr A* 1991;47:392–400.
 54. Kiefer JR, Mao C, Braman JC, Beese LS. Visualizing DNA replication in a catalytically active bacillus DNA polymerase crystal. *Nature* 1998;391:304–307.
 55. Gunasekaran K, Gomathi L, Ramakrishnan C, Chandrasekhar J, Balaram P. Conformational interconversions in peptide β -turns: analysis of turns in proteins and computational estimates of barriers. *J Mol Biol* 1998;284:1505–1516.
 56. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* 1999b;285:1735–1747.
 57. Beaman TW, Sugantino M, Roderick SL. Structure of the hexapeptide xenobiotic acetyltransferase from *Pseudomonas aeruginosa*. *Biochemistry* 1998;37:6689–6696.