Published in final edited form as:

J Am Chem Soc. 2012 February 8; 134(5): 2473–2476. doi:10.1021/ja210045s.

Quasiracemic Crystallization as a Tool to Assess the Accommodation of Non-canonical Residues in Native-Like Protein Conformations

David E. Mortenson[†], Kenneth A. Satyshur[§], Ilia A. Guzei[†], Prof. Katrina T. Forest[§], and Prof. Samuel H. Gellman[†]

Katrina T. Forest: gellman@chem.wisc.edu; Samuel H. Gellman: forest@bact.wisc.edu

[†]Department of Chemistry, University of Wisconsin-Madison, 1101 University Ave. Madison, WI 53706 (USA)

[§]Department of Bacteriology, University of Wisconsin-Madison, 1550 Linden Drive, Madison, WI 53706 (USA)

Abstract

Quasiracemic crystallization has been used to obtain high-resolution structures of two variants of the villin headpiece subdomain (VHP) that contain a pentafluorophenylalanine (F₅Phe) residue in the hydrophobic core. In each case, the crystal contained the variant constructed from L-amino acids and the native sequence constructed from D-amino acids. We were motivated to undertake these studies by reports that racemic proteins crystallize more readily than homochiral forms, and the prospect that quasiracemic crystallization would enable us to determine whether a polypeptide containing a non-canonical residue can closely mimic the tertiary structure of the native sequence. The results suggest that quasiracemic crystallization may prove to be generally useful for assessing mimicry of naturally evolved protein folding patterns by polypeptides that contain unnatural side chain or backbone subunits.

Keywords

Racemic Protein Crystallization; Quasiracemic Proteins; Unnatural Amino Acids; Pentafluorophenylalanine; Villin Headpiece Subdomain

Artificial proteins that retain desirable activities of natural prototypes while displaying improved properties, such as greater conformational or metabolic stability, are subjects of considerable interest. Unnatural subunits, featuring non-canonical side chains and/or backbones, are potentially valuable tools for generating functional mimics of natural proteins.^{1–3} Rational polypeptide engineering with non-canonical subunits, however, requires that one understand the structural consequences of a given substitution. Functions such as recognition and catalysis often depend sensitively on protein conformation, and replacing canonical with non-canonical residues could lead to structural perturbations. The extent of possible perturbation is difficult to predict: thousands of protein structures containing exclusively natural subunits have been determined via X-ray crystallography, but

Correspondence to: Katrina T. Forest, gellman@chem.wisc.edu; Samuel H. Gellman, forest@bact.wisc.edu.

Supporting Information. Experimental details including polypeptide characterization, X-ray diffraction data collection and processing, structural solution and refinement, and calculated RMSD values are available free of charge *via* the Internet at <http://pubs.acs.org>. Model coordinates and structure factors have been deposited in the Protein Data Bank with accession codes: 3TRW (*Pf* racemate); 3TRY (*A* c2 racemate); 3TJW (Phe10 → F₅Phe quasiracemate); 3TRV (Phe17 → F₅Phe quasiracemate).

only a handful of high-resolution data sets are available to show how unnatural subunits are accommodated within protein tertiary structures. Here we illustrate the use of quasiracemic crystallization as a tool for identifying polypeptides that contain unnatural α -amino acid residues yet conformationally mimic a target protein comprising only canonical residues. We have applied this tool to polypeptides containing fluorinated side chains, which have attracted interest because of their unique physical properties and may serve as useful building blocks for rational protein design.²

Racemic pairs of small molecules appear to crystallize more readily than individual enantiomers, although the origin of this trend has been a subject of debate.⁴ Initial studies of racemic protein crystallization were inspired by the advantages of centrosymmetric molecular arrangements (impossible for single enantiomers) for structure solution,⁵ but this motivation has become less urgent with the advent of multiwavelength anomalous dispersion methods for structure determination. Such efforts have been motivated also by the prospect that racemic proteins would crystallize more readily than pure enantiomers,^{5b,e,6} as predicted by Wukovitz and Yeates.⁷ Several examples have shown, however, that racemic proteins need not adopt arrangements obeying rigorous crystallographic symmetry operations in order to cocrystallize.^{6c,8} In the context of these observations, it is intriguing that Kent *et al.* have described one example in which *quasiracemic* protein crystallization was used for heavy atom incorporation into a crystal.^{6a} In a quasiracemic pair of molecules, one partner is very similar to the enantiomer of the other. Pasteur first noted that such pairs could co-crystallize (e.g., (+)-bitartrate and (-)-bimalate salts),⁹ and the phenomenon has subsequently been documented extensively among small molecules.^{9b,10} We were intrigued by the prospect that quasiracemic protein crystallization should provide a sensitive test for the ability of a polypeptide bearing at least one non-canonical subunit to mimic the three-dimensional shape of a prototype containing only canonical residues.

As an initial test of our hypothesis, we have examined variants of the 35-residue villin headpiece subdomain (VHP) that contain pentafluorophenylalanine (F₅Phe). The hydrophobic core of native VHP contains three closely packed Phe side chains (from residues 6, 10 and 17; Figure 1). Previously we used this system to evaluate the effect on conformational stability of Phe $\rightarrow$ F₅Phe substitutions at these core sites.¹¹ Phe10 $\rightarrow$ F₅Phe was found to be moderately stabilizing. NMR analysis of this variant indicated that the spatial arrangement of the three aromatic rings from residues 6, 10 and 17 does not change substantially as a result of incorporating the unnatural side chain.¹² This result suggests that the Phe10 $\rightarrow$ F₅Phe variant constructed from L-residues and VHP itself constructed from D-residues (D-VHP) might form a crystalline quasiracemate, and that attempts to co-crystallize D-VHP with each of the six other L-Phe $\rightarrow$ F₅Phe variants involving positions 6, 10 and 17 would indicate whether any among these polypeptides is a close structural mimic of VHP.

Our crystallization efforts were based on the chicken VHP sequence with a single residue change (Asn27 $\rightarrow$ His; Figure 1), because this variant has provided the highest-resolution VHP crystal structure reported to date (PDB: 1YRF).¹³ In preliminary efforts, two crystalline forms of racemic VHP were obtained, and both structures were solved.¹⁴ One racemate structure proved to be centrosymmetric (space group $P\bar{1}$), while the other is not but contains a 4-fold rotoinversion axis (space group $\bar{4}c2$). Crystallization in space group $P\bar{1}$ was expected based on theoretical arguments,⁷ as well as the high frequency with which this space group has been observed among previously reported racemic protein structures.^{6a,c,d,8a,b} This structure was refined at 2.1 Å resolution to $R_{\text{work}}/R_{\text{free}}$ of 0.291/0.352; these values correspond to $R_{\text{work}}/R_{\text{free}}$ of 0.197/0.241 for a noncentrosymmetric model having similar coordinate error.¹⁵ The structure of the VHP racemate in space group $\bar{4}c2$ was refined at 2.3 Å resolution. Space group $\bar{4}c2$ has not

been previously observed in any protein structure deposited in the Protein Data Bank, but is theoretically possible for co-crystals of enantiomeric molecules.

Among the seven variants of L-VHP generated by replacing one, two or three among Phe6, Phe10 and Phe17 with F₅Phe, two yielded crystals as quasiracemates with D-VHP (Figures 2 and 3). One crystalline quasiracemate involved L-Phe10 → F₅Phe, as anticipated from NMR results,¹² and the other involved L-Phe17 → F₅Phe. In both cases, microcrystals were obtained from ~20% of the precipitant conditions examined in sparse-matrix screening, while the other five quasiracemic mixtures did not produce crystals under any of the 242 conditions examined. Although it is perilous to try to interpret crystallization failures, one is tempted to speculate that the five quasiracemic mixtures that did not produce crystals contained Phe → F₅Phe substitution patterns resulting in conformational instability or a different folding pattern relative to VHP itself.

For the L-(Phe17 → F₅Phe) + D-VHP mixture, the asymmetric unit (i.e., the full unit cell) contains one molecule of each polypeptide, and these quasienantiomers are related by a pseudo-inversion center. This structure, solved by direct methods and refined at 1.00 Å resolution,¹⁴ is essentially pseudo- $P\bar{1}$; however, for the purpose of making comparisons between quasienantiomers, which must be chemically distinct, we refined this structure in the space group P1. Although the arrangement of quasienantiomeric molecules in this structure is pseudocentrosymmetric, the crystal packing is not related in any significant way to the structure of the VHP racemate in space group $P\bar{1}$. The L-(Phe10 → F₅Phe)VHP + D-VHP structure contains a pseudo-4-fold rotoinversion, i.e., this structure can be seen as pseudo- $\bar{4}c2$, but space group F222 was found to give better data processing and refinement statistics than the higher symmetry $\bar{4}c2$, presumably because F222 accommodates the intrinsic differences between quasienantiomers. This structure was refined at 1.46 Å resolution. The asymmetric unit contains one copy each of L-(Phe10 → F₅Phe)VHP and D-VHP (Figure 2). The packing of L- and D-proteins in this structure is virtually identical to that in the $\bar{4}c2$ racemate structure, which contains a crystallographic four-fold rotoinversion operation. The root mean square deviation (RMSD) calculated between *all heavy atoms* in the Phe10 → F₅Phe quasiracemate asymmetric unit (i.e., L-(Phe10 → F₅Phe) + D-VHP) and the corresponding pair of polypeptides in the $\bar{4}c2$ racemate structure (one being generated *via* space group symmetry) is 0.87 Å; this finding indicates that the packing of the quasiracemic pair mimics the packing in the $\bar{4}c2$ racemate structure.

The four new structures we solved provide seven versions of VHP that can be compared with one another and with previously reported structure 1YRF: two independent molecules in the $P\bar{1}$ racemate unit cell, one molecule from the $\bar{4}c2$ structure of racemic VHP, and two each from the quasiracemate structures (the mirror image was used for the D-VHP molecules in these cases). RMSD values were calculated between all pairs of new structures and pairings with 1YRF using various combinations of backbone atoms or only the side chain carbons of the three Phe/F₅Phe residues at positions 6, 10 and 17.¹⁴ The conclusions discussed here are based on the side chain comparisons (Table 1), but the same conclusions can be drawn from the backbone comparisons. Some comparisons between polypeptides from different crystals yielded RMSD values that are less than or comparable to the experimental uncertainty, but in many cases the RMSD values are significant. For both L-(Phe10 → F₅Phe)VHP and L-(Phe17 → F₅Phe)VHP, comparison of the variant with the quasienantiomer in the same crystal provides a measure of the perturbation caused by the non-canonical side chain in very similar packing environments. The RMSD values are ~0.2 Å in both cases, which are larger than estimated coordinate errors but small relative to values calculated between models determined from *different* crystals (Table 1; Table S1). This result indicates that the F₅Phe side chain may effectively mimic the Phe side chain it

replaced. Similar conclusions may be drawn from the relatively small RMSD values (< 0.5 Å) observed for most other comparisons in Table 1.

Somewhat larger differences (RMSD 1.3-1.5 Å) are observed for phenylalanine clusters between either molecule in the Phe17 $\rightarrow$ F₅Phe quasiracemate crystal and all other VHP structures, including 1YRF. The contrast between the similarity of the L-(Phe17 $\rightarrow$ F₅Phe)VHP variant to the mirror image of D-VHP in the same crystal and the more pronounced difference between either member of this pair and the other VHP structures presumably reflects the fact that these polypeptides retain some conformational flexibility in their folded states, which enables structural adaptation between members of a quasiracemic pair upon crystallization. Thus, D-VHP can adopt tertiary structures in which the aromatic side chain cluster RMSD values differ by as much as 1.5 Å in order to adapt to the L-(Phe10 $\rightarrow$ F₅Phe)VHP variant *vs.* the L-(Phe17 $\rightarrow$ F₅Phe)VHP variant.

Both structures of F₅Phe-containing VHP quasiracemates involve pseudo-inversion operations; this result was expected for peptides that are very nearly enantiomeric, based on the predictions of Wukovitz and Yeates.⁷ The structure solved in the space group F222 (Phe10 $\rightarrow$ F₅Phe quasiracemate) does not contain a crystallographic inversion center or rotoinversion axis; however, the packing arrangement of L- and D-peptides in this structure is virtually identical to that of the $\bar{4}c2$ racemate structure, which contains a crystallographic 4-fold rotoinversion operation (Figure 4). Further details of the relationship between these two structures are provided on page S16 of the Supporting Information.

We find that quasiracemic crystals are obtained preferentially over crystals containing only one enantiomer, even though quasienantiomeric peptides cannot crystallize in centrosymmetric arrangements. These observations support the notion that there could be considerable leeway in the selection of peptides that behave as quasienantiomers. Our results provide support for the hypothesis that quasiracemic protein crystallization can be used to identify polypeptides that contain non-canonical residues but nevertheless mimic the tertiary structure of a prototype derived exclusively from natural residues. We have explored this hypothesis in the context of amino acid residues bearing fluorinated side chains, an increasingly popular class of noncanonical building blocks for protein design,² and our efforts have provided the first high-resolution structures in which F₅Phe is accommodated by a folded protein core. It will now be important to determine whether the quasiracemate crystallization strategy can be applied to a broader range of unnatural subunits, including those that differ in the backbone³ rather than in the side chain relative to canonical α -amino acid residues.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the NIH (GM061238, S.H.G.; GM59271, K.T.F.). D.E.M. was supported in part by a Molecular Biophysics Training Grant (T32 GM08293). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor for the support of this research program (Grant 085P1000817). We thank Dr. David Smith and Dr. Zdzislaw Wawrzak of LS-CAT for lending their expertise with data collection, and Mr. David Dyer for many helpful discussions.

References

1. For selected examples of polypeptides containing unnatural sidechains, see: Saks ME, Sampson JR, Nowak MW, Kearney PC, Du F, Abelson JN, Lester HA, Dougherty DA. *J Biol Chem.* 1996; 271:23169. [PubMed: 8798511] Schnarr NA, Kennan AJ. *J Am Chem Soc.* 2001; 123:11081. [PubMed: 11686721] Kirshenbaum K, Carrico IS, Tirrell DA. *Chem Bio Chem.* 2002; 3:2188. Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang Z, Schultz PG. *Science.* 2003; 301:964. [PubMed: 12920298] Zhang Z, Alfonta L, Tian F, Bursulaya B, Uryu S, King DS, Schultz PG. *Proc Nat Acad Sci U S A.* 2004; 101:8882. Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM. *Proc Nat Acad Sci U S A.* 2006; 103:9482. Rodriguez EA, Lester HA, Dougherty DA. *Proc Nat Acad Sci U S A.* 2006; 103:8650. Vogel EM, Imperiali B. *Protein Sci.* 2007; 16:550. [PubMed: 17242376] Ngo JT, Tirrell DA. *Acc Chem Res.* 2011; 44:677. [PubMed: 21815659]
2. For selected examples of polypeptides containing residues with fluorinated side chains, see: Bilgiçer B, Fichera A, Kumar K. *J Am Chem Soc.* 2001; 123:4393. [PubMed: 11457223] Tang Y, Ghirlanda G, Vaidehi N, Kua J, Mainz DT, Goddard WA III, DeGrado WF, Tirrell DA. *Biochemistry.* 2001; 40:2790. [PubMed: 11258889] Tang Y, Ghirlanda G, Petka WA, Nakajima T, DeGrado WF, Tirrell DA. *Angew Chem Int Ed.* 2001; 40:1494. Hodges JA, Raines RT. *J Am Chem Soc.* 2003; 125:9263. Montclare JK, Son S, Clark GA, Kumar K, Tirrell DA. *Chem Bio Chem.* 2009; 10:84. Buer BC, Chugg J, Al Hashimi HM, Marsh ENG. *Biochemistry.* 2010; 49:5760. [PubMed: 20527804] Zheng H, Gao J. *Angew Chem Int Ed.* 2010; 49:8635. Wang F, Qin L, Wong P, Gao J. *Org Lett.* 2011; 13:236. [PubMed: 21142016] Pace CJ, Zheng H, Mylvaganam R, Kim D, Gao J. *Angew Chem Int Ed.* 2012; 51:103.. For a review of fluorinated polypeptides see: Jäckel C, Koksche B. *Eur J Org Chem.* 2005:4483.
3. For selected examples of altered polypeptide backbones, see: Arnold U, Hinderaker MP, Nilsson BL, Huck BR, Gellman SH, Raines RT. *J Am Chem Soc.* 2002; 124:8522. [PubMed: 12121081] Kritzer JA, Lear JD, Hodsdon ME, Schepartz A. *J Am Chem Soc.* 2004; 126:9468. [PubMed: 15291512] Deechongkit S, Nguyen H, Powers ET, Dawson PE, Gruebele M, Kelly JW. *Nature.* 2004; 430:101. [PubMed: 15229605] Jenkins CL, Vasbinder MM, Miller SJ, Raines RT. *Org Lett.* 2005; 7:2619. [PubMed: 15957905] Fu Y, Bieschke J, Kelly JW. *J Am Chem Soc.* 2005; 127:15366. [PubMed: 16262389] Horne WS, Price JL, Keck JL, Gellman SH. *J Am Chem Soc.* 2007; 129:4178. [PubMed: 17362016] Horne WS, Boersma MD, Windsor MA, Gellman SH. *Angew Chem Int Ed.* 2008; 47:2853. David R, Günther R, Baumann L, Lühmann T, Seebach D, Hoffman HJ, Beck-Sickinger AG. *J Am Chem Soc.* 2008; 130:15311. [PubMed: 18942784] Shoulders MD, Satyshur KA, Forest KT, Raines RT. *Proc Nat Acad Sci U S A.* 2010; 107:559.
4. Brock CP, Schweizer WB, Dunitz JD. *J Am Chem Soc.* 1991; 113:9811.
5. (a) Mackay AL. *Nature.* 1989; 342:133.(b) Zawadzke LE, Berg JM. *J Am Chem Soc.* 1992; 114:4003.(c) Zawadzke LE, Berg JM. *Prot Struct Func Gen.* 1993; 16:301.(d) Milton RC, Milton SC, Kent SBH. *Science.* 1992; 256:1445. [PubMed: 1604320] (e) Toniolo C, Peggion C, Crisma M, Formaggio F, Shui X, Eggleston DS. *Nat Struct Biol.* 1994; 1:908. [PubMed: 7773780]
6. (a) Pentelute BL, Gates ZP, Tereshko V, Dashnau J, Vanderkooi JM, Kossiakoff AA, Kent SBH. *J Am Chem Soc.* 2008; 130:9695. [PubMed: 18598029] (b) Mandal K, Pentelute BL, Tereshko V, Kossiakoff AA, Kent SBH. *J Am Chem Soc.* 2009; 131:1362. [PubMed: 19133782] (c) Mandal K, Pentelute BL, Tereshko V, Kossiakoff AA, Kent SBH. *Protein Sci.* 2009; 8:1146. [PubMed: 19472324] (d) Pentelute BL, Mandal K, Gates ZP, Sawaya MR, Yeates TO, Kent SBH. *Chem Comm.* 2010; 46:8174. [PubMed: 20877851] (e) Banigan JR, Mandal K, Sawaya MR, Thammavongsa V, Hendrickx AP, Schneewind O, Yeates TO, Kent SBH. *Protein Sci.* 2010; 19:1840. [PubMed: 20669184]
7. Wukovitz SW, Yeates TO. *Nat Struct Biol.* 1995; 2:1062. [PubMed: 8846217]
8. (a) Patterson WR, Anderson DH, DeGrado WF, Cascio D, Eisenberg D. *Prot Sci.* 1999; 8:1410.(b) Hung LW, Kohmura M, Ariyoshi Y, Kim SH. *J Mol Biol.* 1999; 285:311. [PubMed: 9878408]
9. (a) Pasteur L. *Ann Chim Phys.* 1853; 28:437.(b) Wheeler KA, Grove RC, Davis RE, Kassel WS. *Angew Chem Int Ed.* 2008; 47:78.
10. (a) Karle IL, Karle J. *J Am Chem Soc.* 1966; 88:24.(b) Lineberry AM, Benjamin ET, Davis RE, Kassel WS, Wheeler KA. *Cryst Growth Des.* 2008; 8:612.(c) Breen ME, Tameze SL, Dougherty WG, Kassel WS, Wheeler KA. *Cryst Growth Des.* 2008; 8:3863.(d) Grover RC, Malehorn SH,

- Breen ME, Wheeler KA. Chem Comm. 2010; 46:7322. [PubMed: 20835474] (e) Kelley SP, Fábíán L, Brock CP. Acta Cryst. 2011; B67:79.
11. Woll MG, Hadley EB, Mecozzi S, Gellman SH. J Am Chem Soc. 2006; 128:15932. [PubMed: 17165695] . For extension of this approach to include partially fluorinated Phe derivatives, see: Zheng H, Comeforo K, Gao J. J Am Chem Soc. 2009; 131:18. [PubMed: 19067509]
12. Cornilescu G, Hadley EB, Woll MG, Markley JL, Gellman SH, Cornilescu CC. Prot Sci. 2007; 16:14.
13. Chiu TK, Kubelka J, Herbst-Irmer R, Eaton WA, Hofrichter J, Davies DR. Proc Nat Acad Sci U S A. 2005; 102:7517.
14. See Supporting information for tables of RMSD values calculated from positions of polypeptide backbone atoms and details of structure solutions.
15. Luzzati PV. Acta Cryst. 1952; 5:802.

Figure 1.

Previously reported structure of a VHP variant (PDB code: 1YRF) and sequence of VHP. Phe → F₅Phe substitutions were made for core phenylalanine residues (shown in red).

Figure 2.

(a) $2F_o - F_c$ and $F_o - F_c$ maps of density surrounding F₅Phe10, contoured at 1.5σ and $\pm 3.5\sigma$, respectively. (b) Structure of L-(Phe10 → F₅Phe)VHP + D-VHP quasiracemate.

Figure 3.

(a) Structure of (L-Phe10 → F₅Phe)VHP obtained *via* quasiracemic crystallization. (b) Structure of (L-Phe17 → F₅Phe)VHP obtained *via* quasiracemic crystallization. (c) Overlay of the two F₅Phe-containing forms of VHP with L-VHP from the racemate structure obtained in space group $\bar{A}c2$. Side chain carbon atoms are shown in white, with fluorine atoms in F₅Phe residues shown in green.

Figure 4.

(a) Crystallographic unit cell of the VHP racemate structure in space group $\bar{A}c2$, with polypeptides from adjacent cells shown for comparison to the quasiracemic structure at right. The unit cell itself contains eight copies each of L-VHP (magenta) and D-VHP (yellow). (b) Crystallographic unit cell of the (Phe10 $\rightarrow$ F₅Phe) quasiracemate structure in space group F222. The unit cell contains sixteen copies each of L(Phe10 $\rightarrow$ F₅Phe)VHP (red) and D-VHP (yellow).

Table 1

^a

	<i>P</i> I racemate		<i>I</i> 4c2 racemate		Phe10 → F ₅ Phe		Phe17 → F ₅ Phe	
	L ₁	L ₂	D	L	D	L	D	
<i>P</i> I racemate	L ₁	-						
	L ₂	0.15	-					
<i>I</i> 4c2 racemate	D	0.29	0.28	-				
	L	0.32	0.35	0.22	-			
Phe10 → F ₅ Phe	D	0.26	0.28	0.11	0.20	-		
	L	1.34	1.32	1.48	1.46	1.46	-	
Phe17 → F ₅ Phe	D	1.26	1.24	1.39	1.38	1.38	0.18	
	L	0.29	0.26	0.43	0.45	0.46	1.37	
1YRF							1.28	

^aRMSD values (Å) calculated between all pairs of structures under study and PDB structure 1YRF, using core Phe and F₅Phe side chain atoms. The RMSD values calculated between polypeptides within a crystal are small (0.15 – 0.20 Å) while RMSD values calculated between polypeptides in different crystals vary considerably. Values were calculated using the positions of all 21 side chain carbon atoms (Cβ, Cγ, Cδ1+2, Cε1+2, Cζ); fluorine atoms were not included in these calculations. Those values in italics are lower than estimated uncertainties in atomic positions obtained from refined models.