PROTEIN STRUCTURE REPORT

Dioxane contributes to the altered conformation and oligomerization state of a designed engrailed homeodomain variant

GEOFFREY K. HOM,¹ J. KYLE LASSILA,¹ LEONARD M. THOMAS,^{2,3} AND
STEPHEN L. MAYO^{2,3,4}

¹Biochemistry and Molecular Biophysics Option, ²Howard Hughes Medical Institute, ³Division of Biology and Chemical Engineering, and ⁴Division of Chemistry, California Institute of Technology, Pasadena, California 91125, USA

(RECEIVED December 8, 2004; FINAL REVISION December 16, 2004; ACCEPTED December 20, 2004)

Abstract

Our goal was to compute a stable, full-sequence design of the *Drosophila melanogaster* engrailed homeodomain. Thermal and chemical denaturation data indicated the design was significantly more stable than was the wild-type protein. The data were also nearly identical to those for a similar, later full-sequence design, which was shown by NMR to adopt the homeodomain fold: a three-helix, globular monomer. However, a 1.65 Å crystal structure of the design described here turned out to be of a completely different fold: a four-helix, rodlike tetramer. The crystallization conditions included ~25% dioxane, and subsequent experiments by circular dichroism and sedimentation velocity analytical ultracentrifugation indicated that dioxane increases the helicity and oligomerization state of the designed protein. We attribute at least part of the discrepancy between the target fold and the crystal structure to the presence of a high concentration of dioxane.

Keywords: protein design; dioxane; engrailed homeodomain

The original purpose of this project was to computationally design an amino acid sequence that stably adopts the homeodomain fold. The target fold was the same as for previous homeodomain designs from our lab (Marshall and Mayo 2001; Marshall et al. 2002): a 51-residue, crystallographically well-defined fragment of the *Drosophila melanogaster* engrailed homeodomain (Fig. 1A; Clarke et al. 1994). This fragment is a globular, three-helix monomer. As in our previous homeodomain designs, we did not consider DNA binding but rather focused on protein stability.

We designed two sequences, UMC and UVF (P.S. Shah, G.K. Hom, S.A. Ross, and S.L. Mayo, in prep.). UMC was obtained via a Monte Carlo algorithm (Voigt et al. 2000).

UVF has a slightly lower computed energy and could be obtained via either the Vegas (Shah et al. 2004) or the FASTER (Desmet et al. 2002) algorithm.

UMC and UVF have 79% sequence identity and also have nearly identical thermal and chemical denaturation profiles. For both proteins, the melting temperature is >99°C and $\Delta G_{\text{unfolding}}$ is 4.2 kcal/mol. The one-dimensional ¹H NMR spectra of the proteins display the characteristics expected of well-folded proteins, and the NMR-determined structure of UVF matches the homeodomain fold (P.S. Shah, G.K. Hom, S.A. Ross, and S.L. Mayo, in prep.).

The above evidence indicated that UMC also adopts the homeodomain fold. However, a crystal structure of UMC would give direct confirmation of the overall fold and allow for a detailed comparison of crystallographic and computed side-chain conformations, which would provide critical data for improving our protein design algorithm (Dahiyat and Mayo 1996).

Reprint requests to: Stephen L. Mayo, Professor of Biology and Chemistry, Caltech/HHMI 114-96, 1200 E. California Boulevard, Pasadena, CA 91125-9600, USA; e-mail: steve@mayo.caltech.edu; fax: (626) 568-0934.

Article published online ahead of print. Article and publication date are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.041277305>.

Figure 1. (A) Target homeodomain fold for UMC. (B) Ribbon diagram of the UMC crystal structure. The coloring is cyan (chain A), green (chain B), brown (chain C), salmon (chain D), and purple (cadmium). (C) Coordination of one of the two cadmium atoms by four glutamates. The coloring is cyan (chain A), green (chain B), yellow (chain C), purple (cadmium), and red (oxygen). The σ_A -weighted density map is contoured at 2σ , up to 3.5 Å from the cadmium. Chain C is from a symmetry-related molecule of that shown in B. (D) Dioxane molecules mediating helix-helix packing. The coloring is green (chain B), brown (chain C), yellow (dioxane), red (oxygen), and blue (nitrogen). The σ_A -weighted density map is contoured at 1σ , up to 3 Å from the dioxane. Figures were generated in PyMOL (<http://www.pymol.org>).

Here we report a 1.65 Å crystal structure of UMC. The structure is a rodlike, four-helix tetramer (Fig. 1B), not the expected globular, three-helix monomer. This discrepancy could be due to a lack of explicit negative design in our design algorithm; however, because of the similarity of UMC to the successful UVF design, we investigated if the crystallographic conditions could be responsible for the discrepancy. In particular, the role of dioxane was examined.

Results

The crystal structure of UMC was determined by using single wavelength anomalous diffraction (SAD). Crystallographic statistics are shown in Table 1. The asymmetric unit contains four UMC molecules forming an antiparallel helical bundle with one UMC molecule per helix. Main-chain and side-chain density could not be interpreted for some terminal residues (residues 1–3 of chain A; 1–4, 51–52 of chain B; 1–2, 47–52 of chain C; and 1–4, 51–52 of chain D). The asymmetric unit also contains 2 cadmium atoms, 1 acetate molecule, and 10 dioxane molecules. The cadmium atoms are each coordinated by four carboxylate anions: one cadmium is coordinated by four glutamate side-chains (Fig.

1C), and the other cadmium is coordinated by three glutamate side-chains and an acetate molecule.

Several dioxane molecules mediate helix-helix packing (Fig. 1D). This observation led us to examine the effect of dioxane on the helicity and oligomerization of UMC in solution.

Helicity was examined by far-UV circular dichroism. Ellipticity was virtually unchanged when UMC was exposed to CdCl_2 alone (data not shown) or CdCl_2 with 10% dioxane (Fig. 2A). However, exposure of UMC to 20% dioxane lowered the minima at 208 and 222 nm and was thus indicative of an increase in helicity. The increase, while significant, was still less than that for 30% trifluoroethanol (TFE), a helix stabilizer (Rohl et al. 1996). Higher percentages of dioxane did not further increase the helicity significantly (data not shown).

Oligomerization was examined by sedimentation velocity analytical ultracentrifugation. Exposure to 20% dioxane significantly decreased the percentage of monomeric UMC, from 81.4% to 62.8%, and concomitantly increased the percentage of dimeric UMC, from 14.8% to 36.3% (Fig. 2B). The frictional ratio, which describes the shape of the sedimenting species, also increased. A sphere has a ratio of

Table 1. X-ray data collection and refinement statistics

	R-axis IV	SSRL ^a
Unit cell		
<i>a</i>	50.767 Å	50.712 Å
<i>b</i>	52.562 Å	52.646 Å
<i>c</i>	82.147 Å	82.182 Å
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Wavelength	1.5418 Å	0.8265 Å
Resolution range	81.65–1.90 Å	44.32–1.65 Å
No. of reflections collected	208,991	204,302
No. of unique reflections	17,953	27,060
R _{merge} ^b	5.6% (55.1%) ^c	4.7% (19.6%)
I/σ(I)	10.1 (1.3)	31.8 (8.5)
Completeness	99.9% (99.5%)	99.8% (100.0%)
Final refinement		
R _{cryst}		18.7%
R _{free} ^d		22.7%
Figure of merit		0.863
No. of residues		368
No. of water molecules		168
No. of non protein molecules		11
Mean B value		28.1 Å ²
RMSD from standard stereochemistry		
Bond length		0.017 Å
Bond angle		1.527°
Ramachandran plot statistics		
Most favored regions		99.4%
Additional allowed regions		0.6%
Generously allowed regions		0.0%
Disallowed regions		0.0%

^a Stanford Synchrotron Radiation Laboratory.^b $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$, where *I* is the observed intensity and $\langle I \rangle$ is the average intensity.^c Numbers in parentheses represent values in the highest resolution shell (1.90–1.99 Å for the R-axis IV data and 1.652–1.695 Å for the SSRL data).^d R_{free} was calculated for 5% of randomly selected reflections excluded from refinement.

~1.2, whereas rodlike shapes have higher ratios. The frictional ratio increased from 1.22 to 1.42 in the presence of dioxane.

Discussion

Our crystal structure of UMC is quite dissimilar to the target homeodomain fold. Instead of three short helices, each monomer is a single long helix. However, the crystallization conditions, especially the high concentration of dioxane, may induce UMC into a conformation unrepresentative of UMC in solution.

Increased helicity and oligomerization due to dioxane

Dioxane increased the helicity of UMC. While dioxane had a significant effect, the $[\theta]_{222}$ for 20% dioxane (–25,000 deg

cm²/dmol res) was less negative than for 30% TFE (–31,000 deg cm²/dmol res).

The effect of dioxane on increasing helicity has been reported previously (Tanford et al. 1960; Urnes and Doty 1961; Iizuka and Yang 1965). The increase in helicity can be explained entropically: Nonpolar solvent increases the entropic cost of forming protein hydrogen bonds to water and thus decreases the relative cost of forming helical hydrogen bonds. The use of organic solvents may have played a role in the crystallization of a number of short aminoisobutyric acid-containing peptides, which also adopt extended continuous helical structures (Karle 1992).

The sedimentation velocity data showed that dioxane increases the oligomerization state and frictional ratio of UMC. While there was a significant increase in the amount of dimer, there was no evidence of a tetramer, as might be expected from the crystal structure. One explanation is that formation of a tetrameric species requires cadmium. Al-

Figure 2. (A) Far-UV circular dichroism analysis of UMC. The spectra are UMC (dashes), UMC in 5 mM CdCl₂ and 10% dioxane (triangles), UMC in 20% dioxane (crosses), and UMC in 30% TFE (circles). (B) Molar mass distribution of UMC as determined by sedimentation velocity: UMC (solid line); UMC in 20% dioxane (dotted line).

though CdCl_2 alone and CdCl_2 with 10% dioxane had no effect on the helicity of UMC, low millimolar amounts of CdCl_2 (e.g., 2–5 mM) caused essentially all UMC to precipitate out in the presence of >15% dioxane. The UMC crystals appeared a couple of weeks after precipitate had formed in the well. Perhaps cadmium further increases the dioxane-induced helicity and/or oligomerization of UMC but requires the very slow mixing that occurs in the crystallization well.

Conclusion

Overall, the crystal structure has increased helicity and altered oligomerization compared with the target fold. Both of these differences were inducible by dioxane. We thus attribute at least part of the discrepancy between the target fold and the crystal structure of the designed sequence to the presence of a high concentration of dioxane. Although low concentrations (1%–2%) of dioxane have been reported to improve crystallization of some proteins (e.g., Sigler et al. 1966; Matthews et al. 1967), we suggest that high concentrations of dioxane be used with caution.

Materials and methods

Protein design and purification

The UMC design, construction, expression, and purification were similar to our previous engrailed designs (Marshall and Mayo 2001; Marshall et al. 2002) and are described in detail elsewhere (P.S. Shah, G.K. Hom, S.A. Ross, and S.L. Mayo, in prep.). A brief summary is below.

The starting model for all engrailed designs was Protein Data Bank (PDB) entry 1enh (Clarke et al. 1994). Because residue 35 has a positive ϕ angle, it was preserved as glycine. The UMC design protocol was identical to the B6 design protocol of Marshall and Mayo (2001), except that in the UMC design all residues were designed simultaneously, and a Monte Carlo simulation (Voigt et al. 2000) was used instead of a dead-end elimination–based algorithm (Gordon et al. 2003) to find a low-energy sequence. The protein was expressed in *Escherichia coli* and purified via freeze-thaw (Johnson and Hecht 1994) followed by HPLC using an acetonitrile/water gradient containing 0.1% TFA. Mass spectrometry indicated UMC has an N-terminal methionine.

Crystallization

Crystals were obtained by using a modified sitting drop method that utilizes a “reservoir mimic.” The well reservoir is minimized to contain only the volatile reagents and NaCl. The nonvolatile reagents normally in the reservoir are kept in a separate solution (the mimic) that is only added to the crystallization drop.

We also used Fluorinert (Hampton Research), which is expected to be denser than the drop and to allow the drop to float. Under our conditions, Fluorinert floated above the drop. However, this serendipitously slowed the otherwise rapid crystal degradation that would happen upon well opening and presumably due to the volatility of dioxane.

The initial crystallization condition was 35% dioxane (Hampton Research Crystal Screen 2). The final crystallization conditions were as follows: the well reservoir contained 500 μL of either 24% or 25% dioxane; the well post contained 1 μL of protein solution (~17 mg/mL UMC, 50 μM sodium citrate at pH 5.5) followed by 1 μL of reservoir mimic solution (0.1 M MES at pH 5.7, 30% PEG 400, 10–15 mM CdCl_2); and 20 μL of Fluorinert (Hampton Research) was then added on top of each post. Trays were incubated at 20°C; crystals appeared after about 2 wk. The largest crystals had dimensions of $\sim 150 \times 150 \times 200 \mu\text{m}$.

Structure determination

Data were collected by using a Cu source on a Rigaku RU3HR generator with an R-Axis IV detector at 100 K. Data were processed by using the HKL program suite v1.97.9 (Otwinowski and Minor 1997). Initial electron density maps were generated by using SAD phasing as implemented in the program suite *Elves* (Holton and Alber 2004). The final model was determined by subsequent rounds of building and refinement using *O* (Jones et al. 1991) and *REFMAC* (Murshudov et al. 1999) from the CCP4 program suite (Collaborative Computational Project 1994) to an R-factor of 22.2% ($R_{\text{free}} = 27.8\%$). Final refinement was done with high resolution data collected at beamline 9.2 at the Stanford Synchrotron Radiation Laboratory and produced a final R-factor of 18.7% ($R_{\text{free}} = 22.7\%$).

Coordinates and structure factors have been deposited in the PDB under the accession code 1Y66.

Circular dichroism and sedimentation velocity

Circular dichroism data were collected on an Aviv 62DS spectrometer equipped with a thermoelectric unit. Wavelength scans were done from 190–250 nm at 20°C in a 0.1-mm path-length cell. All samples contained 532 μM UMC and 10 mM sodium citrate (pH 5.5). Protein concentration was determined by absorbance at 280 nm in the presence of 8 M guanidine HCl.

Sedimentation velocity data were collected on a Beckman XL-I analytical ultracentrifuge with interference optics. Samples contained 532 μM UMC and 0.1 M sodium citrate (pH 5.5). Samples were dialyzed for 3 h at room temperature against ~100 mL of the corresponding solution without protein. A 12-mm Epon centerpiece and sapphire windows were used. The rotor, an An-60 Ti, was spun at 55,000 rpm at 25°C. Scans were taken every 5 min for ~15 h. Data were analyzed with SEDFIT (Schuck 2000).

Acknowledgments

We are grateful to Premal Shah, Rhonda Digiusto, Scott Ross, and Karin Crowhurst for assistance with NMR; Doug Rees, James Holton, and J.J. Plecs for assistance with crystallography; Po-Ssu Huang for assistance with sedimentation velocity experiments and PyMOL; and Marie Ary and Jessica Mao for assistance with the manuscript. This work was supported by the Howard Hughes Medical Institute, the Ralph M. Parsons Foundation, the Defense Advanced Research Projects Agency, and an IBM Shared University Research Grant. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory (SSRL), a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental

Research, and by the NIH, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences. We thank the Gordon and Betty Moore Foundation for their support of the crystallographic resources of the Molecular Observatory for Structural Molecular Biology used in this study.

References

- Clarke, N.D., Kissinger, C.R., Desjarlais, J., Gilliland, G.L., and Pabo, C.O. 1994. Structural studies of the engrailed homeodomain. *Protein Sci.* **3**: 1779–1787.
- Collaborative Computational Project. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**: 760–763.
- Dahiyat, B.I., and Mayo, S.L. 1996. Protein design automation. *Protein Sci.* **5**: 895–903.
- Desmet, J., Spriet, J., and Lasters, I. 2002. Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization. *Proteins* **48**: 31–43.
- Gordon, D.B., Hom, G.K., Mayo, S.L., and Pierce, N.A. 2003. Exact rotamer optimization for protein design. *J. Comput. Chem.* **24**: 232–243.
- Holton, J. and Alber, T. 2004. Automated protein crystal structure determination using ELVES. *Proc. Natl. Acad. Sci.* **101**: 1537–1542.
- Iizuka, E. and Yang, J.T. 1965. Effect of salts and dioxane on the coiled conformation of poly-L-glutamic acid in aqueous solution. *Biochemistry* **4**: 1249–1257.
- Johnson, B.H. and Hecht, M.H. 1994. Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing. *Biotechnology (N Y)* **12**: 1357–1360.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**(Pt. 2): 110–119.
- Karle, I.L. 1992. Folding, aggregation and molecular recognition in peptides. *Acta Crystallogr. B* **48**(Pt. 4): 341–356.
- Marshall, S.A. and Mayo, S.L. 2001. Achieving stability and conformational specificity in designed proteins via binary patterning. *J. Mol. Biol.* **305**: 619–631.
- Marshall, S.A., Morgan, C.S., and Mayo, S.L. 2002. Electrostatics significantly affect the stability of designed homeodomain variants. *J. Mol. Biol.* **316**: 189–199.
- Matthews, B.W., Sigler, P.B., Henderson, R., and Blow, D.M. 1967. Three-dimensional structure of tosyl- α -chymotrypsin. *Nature* **214**: 652–656.
- Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., and Dodson, E.J. 1999. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D Biol. Crystallogr.* **55**(Pt. 1): 247–255.
- Otwinowski, Z. and Minor, W. 1997. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**(Pt. A): 307–326.
- Rohl, C.A., Chakrabarty, A., and Baldwin, R.L. 1996. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* **5**: 2623–2637.
- Schuck, P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**: 1606–1619.
- Shah, P.S., Hom, G.K., and Mayo, S.L. 2004. Preprocessing of rotamers for protein design calculations. *J. Comput. Chem.* **25**: 1797–1800.
- Sigler, P.B., Jeffery, B.A., Matthews, B.W., and Blow, D.M. 1966. An x-ray diffraction study of inhibited derivatives of α -chymotrypsin. *J. Mol. Biol.* **15**: 175–192.
- Tanford, C., De, P.K., and Taggart, V.G. 1960. The role of the α -helix in the structure of proteins. Optical rotatory dispersion of β -lactoglobulin. *J. Am. Chem. Soc.* **82**: 6028–6034.
- Urnes, P. and Doty, P. 1961. Optical rotation and the conformation of polypeptides and proteins. In *Advances in protein chemistry* (eds. C.B. Anfinsen Jr. et al.), pp. 401–543. Academic Press, New York.
- Voigt, C.A., Gordon, D.B., and Mayo, S.L. 2000. Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. *J. Mol. Biol.* **299**: 789–803.