Possible Exceptions to Rules for α-Helix Termination by Glycine

Recently R. Aurora *et al.* (*1*) gave rules to determine, on the basis of a primary amino acid sequence, whether a propagating α helix on reaching a glycine residue will terminate or propagate through the glycine. We have found several possible exceptions to the rules stated in the report by Aurora *et al.* in their initial set of 42 proteins (Table 1) as well as in two other proteins, the x-ray crystal structures of which have been published (*2*, *3*) since the report.

For α helices and flanking residues, we use the following notation (*1*)

. . . –N″–N′–Ncap–N1–N2–N3– . . .

. . . –C3–C2–C1–Ccap–C′–C″– . . .

where the numbered residues belong to the helix, the primed residues belong to the flanking sequences, and Ncap and Ccap belong to both the helix and the adjacent flanking region. Briefly, the rules for helix termination by one stereochemical motif, the Schellman motif, can be summarized as follows (*1*): We temporarily assume that a glycine residue is at C′ and examine the amino acids surrounding the glycine. If the amino acid at C″ (the amino acid after the glycine) is apolar (Ala, Val, Ile, Leu, Met, Phe, Trp, Cys) or Lys or Arg, and the amino acid at C1 is polar (Ser, Thr, Asn, Asp, Gln, Glu, Arg, Lys) or Ala, and at least one of C2, C3, or C4 is apolar or Lys or Arg, then termination by the Schellman motif is predicted. If C1 is apolar (not Ala) or C2, C3, and C4 are all polar (not Lys or Arg), then helix continuation is predicted.

In the initial set of 42 proteins (*1*), there were only five instances reported in which the rules of Aurora *et al.* predicted helix termination at a glycine, but termination did not occur. Three of these five exceptions were a result of ligand binding; in the other two, glycine was at the N5 position of a helix out to which strong NH_2-capping interactions might extend. In another set of 26 proteins (*1*) containing 34 glycines within helices, the rules correctly predicted helix continuation through these glycines in 22 cases. Of the 12 exceptions, 10 were a result of ligand binding.

We have found seven more possible exceptions to the rules of Aurora *et al.* in their initial set of 42 proteins (Table 1). In five instances the rules predict helix termination, but experiment shows helix continuation. In one instance the rules predict helix continuation, but experiment shows helix termination. We also found one instance in which the rules make no prediction. In this case, besides a glycine at C′, there is also a glycine at C1. The rules of Aurora *et al.* are silent as to continuation or termination in the case of glycine at C1.

As many exceptions to the rules of Aurora *et al.* are a result of ligand binding, it might be helpful to be able to predict instances of such binding. We wonder if such prediction is possible. Also, are there examples in which helix termination is found experimentally despite strong NH_2-capping interactions (conversely to the examples of helix continuation perhaps due to NH_2-capping interactions extending out to position N5 of a helix where a helix termination has been predicted)?

In the x-ray crystal structure of the binding domain of methionine synthase (*2*), the rules of Aurora *et al.* predict helix termination by the Schellman motif at G^{865} (ASRTVG^{865}VV), but experiment gives helix continuation. In the crystal structure of the 60K subunit of the nickel-iron hydrogenase from *Desulfovibro gigas* (*3*), the rules of Aurora *et al.* predict helix termination by the Schellman motif at G^{101} (RNLTMG^{101}AQ), G^{409} (LFSTLG^{409}RT), and G^{415} (RTAARG^{415}IQ), but experiment gives helix continuation in all three instances.

If the possible exceptions to the rules of Aurora *et al.* can be understood and predicted, these rules will be of importance in our ability to predict protein tertiary structure from primary sequence data. If not, it might cause us to question the extent to which protein folding is determined by local sequence information.

Note added in proof: Since submitting this comment, we have found another possible exception to the rules of Aurora *et al.*: In the protein glutathione reductase (3grs) for Gly^{346} (VAIAAG^{346}RK), the rules of Aurora *et al.* predict helix termination, but experiment gives helix continuation. Also, the proteins avian pancreatic polypeptide (1ppt) and plastocyanin (2pcy) from the original set of 42 proteins have no helices with internal glycines or helices terminated with glycine at C′ (*1*, *4*).

Eric L. Altschuler
Martin Lades
Institute for Scientific Computing Research, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA

Table 1. Possible exceptions to rules for α-helix termination by glycine in the original set of 42 proteins of Aurora *et al.* (*1*).

Protein (PDB code)	Helix boundaries (Ncap-Ccap) (reference)	Sequence	Prediction	Experiment
2cts	393-415(*1*)	VSRALG^{404}VL	Termination	Continuation
4fxn	10-26 (*1*)	ELIAKG^{22}II	Termination	Continuation
4fxn	124-136(*4*)	DCIEFG^{132}KK	Termination	Continuation
1mbo	58-77 (*4*)	DLKKHG^{55}VT	Termination	Continuation
1mbo	58-77 (*4*)	VLTALG^{63}AI	Termination	Continuation
2cts	37-43 (*1*)	VDMMYG^{44}GM	Continuation*	Termination
4fxn	93-106(*1*)	RMNGYG^{107}CV	—†	Termination

*In the case of Gly-Gly the rules for helix termination by the Schellman motif or helix continuation are applied if C‴ is apolar or Lys or Arg (*1*). †The rules of Aurora *et al.* make no prediction in this case.

REFERENCES AND NOTES

1. R. Aurora, R. Srinivasan, G. D. Rose, *Science* **264**, 1126 (1994).
2. C. L. Drennen *et al.*, *ibid.* **266**, 1669 (1994).
3. A. Volbeda *et al.*, *Nature* **373**, 580 (1995).
4. L. Presta and G. D. Rose, *Science* **240**, 1632 (1988).
5. This work was performed at the Lawrence Livermore National Laboratory under the auspices of the U.S. Department of Energy contract number W-7405-ENG-48.

6 February 1995; revised 12 March 1995; accepted 19 May 1995

Response: We thank Altschuler and Lades for their careful reading of our report and reanalysis of the data. Our work (*1*) emphasized the fact that helices terminate in a limited number of ways, and we described two such motifs, the Schellman and the α_L, in detail. Both are glycine-based structures found frequently at the COOH-termini of protein helices. The patterns observed in these motifs are sufficiently distinct that simple rules can be formulated to predict whether the presence of a glycine residue will cause a helix to terminate at all, and, if so, whether it will terminate in a Schellman or an α_L motif.

The main focus of Altschuler and Lades' analysis is the Schellman motif. In this motif, hydrophobic residues (Lys or Arg act as hydrophobic residues in this case) at C″ (that is, the residue following the glycine) and C3 interact, together with a 6 → 1, 5 → 2 hydrogen bond pattern. The hydrophobic interaction establishes a hydrophobic surface and, if the helix is amphipathic, then C1, which is situated on the opposite helical face, will be solvent-exposed. On

Table 1. Exceptions to rules for α-helix termination by glycine.

Case no.	Protein (PDB code)	Helix boundaries (Ncap-Ccap)	Gly no.	Prediction	Experiment	Explanation
1	2cts	393-415	404	Termination	Continuation	*
2	4fxn	10-26	22	Termination	Continuation	FMN binding
3	4fxn	124-136	132	Termination	Continuation	Salt bridge Lys-Glu
4	1mbo	58-77	65	Termination	Continuation	Heme binding
5	1mbo	58-77	73	Termination	Continuation	Heme binding
6	2cts	37-43	44	Continuation	Termination	Tyr is amphipathic
7	4fxn	93-106	107	None	Termination	Gly is solvent-exposed

*A Schellman motif is formed when C3 interacts with C″. When C3 is polar and the interaction is with C2 or C4 instead, the residue at C″ must have a long side chain (for example, Lys, Arg, Trp, and so forth) that can reach its interaction partner at C2 or C4.

occasion, this simple idea is confounded by additional factors—for example, involvement of metals and prosthetic groups—leading to a "violation" of the rule. This is described in detail in our report.

The polarity of the C1 position is a key factor in our rules. But other positions are also involved, and particular combinations of residues can lead to an energetic "tug of war" in some instances. Rather than provide an exhaustive list of the possibilities, the intent of our report was to focus on the simple underlying ideas and to analyze their validity in monomeric proteins.

Altschuler and Lades raise examples that help clarify several points. They are minor points in our opinion, which is why we did not focus on them in our report.

It is true that our rules, as stated, do not predict case 1 (Table 1) correctly. In our report, we glossed over the fact that, when C3 is polar, side chain length must be assessed to decide whether the apolar residue at C″ can reach an apolar interaction partner at C2 or C4. Such complexities occur in an extremely small fraction of the total cases, but methionine synthase may be another example. Cases 2, 4, and 5 in Table 1 involve a prosthetic group, which can override the rules, as described in our report; we were remiss in not including these examples in our table of exceptions (*1*, table 3). As we noted in table 3, a salt bridge is one of the factors that can promote deviations from the rules, and that is what happens in case 3 (Table 1). Additional context-dependent effects come into play when C1 is amphipathic [as noted in table 2 in (*1*)], and this situation was deliberately excluded from the rules; case 6 (Table 1) is an example. In case 7 (Table 1), termination would place the glycine at the solvent-exposed C1 position, which is entirely compatible with the rules, although not specified explicitly in table 2 in (*1*).

Altschuler and Lades mention more "possible exceptions" in our glycine termination rules, taken from a single new structure, nickel-iron hydrogenase (*2*), a multimer. Our report, submitted for publication in December 1993, was written almost 2 years ago. After publication, the basic ideas presented therein have been affirmed repeatedly, both in proteins and in peptides (*3*). Even the hydrogenase structure (*2*) includes sites that are consistent with the rules (although Altschuler and Lades omit mention of them). At this point, an independent reassessment of our findings using newer data would be welcome. However, valid assessment involves documenting both failures and successes of the rules in a representative set of molecules, not merely dredging the literature for potential counterexamples.

Finally, Altschuler and Lades would like to have further rules to predict ligand binding and NH_2-terminal capping. So would we, but that goal was beyond the scope of our report.

George Rose
Rajeev Aurora
Rajgopal Srinivasan
Department of Biophysics and Biophysical Chemistry, School of Medicine, Johns Hopkins University, 725 North Wolfe Street, Baltimore, MD 21205–2185, USA

REFERENCES

1. R. Aurora, R. Srinivasan, G. D. Rose, *Science* **264**, 1126 (1994).
2. A. Volbeda *et al.*, *Nature* **373**, 580 (1995).
3. Y. Gong, H. X. Zhou, M. Guo, N. R. Kallenbach, *Protein Sci.*, in press.

9 March 1995; accepted 19 May 1995