

# Protein folding and the Paracelsus challenge

George D. Rose

**A challenge to change one protein into another while retaining 50% of the original protein's sequence has been met and provides a warning to other would-be protein folding/engineering challenges: only offer a prize of a tee-shirt.**

A few years ago, Trevor Creamer (Department of Biochemistry, School of Medicine, University of Kentucky) and I formulated the Paracelsus challenge: transform the conformation of one globular protein into that of another by changing no more than half the sequence<sup>1</sup>. A cash award was included as incentive. Now, on page 110 of this issue of *Nature Structural Biology*, Dalal *et al.*<sup>2</sup> describe their success in meeting this challenge, and Lynne Regan is scheduled to accept the award at a Hopkins ceremony on 25 June 1997. Why is theirs such a remarkable achievement?

Like Gulliver, the protein's conformation is determined by a large number of small Lilliputian constraints<sup>3</sup>. To test this view, Creamer and I devised the Paracelsus challenge, named after the

16<sup>th</sup> century Swiss physician, an alchemist who more-or-less invented pharmacology. Frankly, we thought it could not be done. Putting our money where our mouths were, we put forth the challenge, with arbitration provided by Eaton Lattman, editor-in-chief of *Proteins, Structure Function and Genetics* and chairman of the Jenkins Department of Biophysics at Johns Hopkins. As a reward, we included a \$1,000 prize, a sum reckoned as the smallest amount that might be noteworthy and the largest we could hope to scrape together. Would that it had been a tee-shirt. Dalal *et al.*<sup>2</sup> succeeded by transforming a predominantly  $\beta$ -sheet protein into an all  $\alpha$ -helix protein. Still, many aspects of the challenge persist, as described below, and virtue will be its own reward for future

participants — no mere platitude as far as we are concerned.

How did Dalal *et al.* tackle the problem, and what design lessons can be learned from their approach? Their procedure was straightforward: identify two proteins of similar size but dissimilar structure and transform one into the other, initial  $\rightarrow$  final. In practice, this strategy is more stringent than required by the challenge. While a suitably chosen final structure is surely a useful guide, the actual product structure need only be globular and folded, with a different conformation. Here, protein G (the B1 domain of *Streptococcal* IgG-binding protein) was chosen as the initial protein and Rop (repressor of primer, a transcriptional regulator) as the final one. Each structure is known<sup>4,5</sup> (see Fig. 1 in

Dalal *et al.*<sup>2</sup>), and the Regan laboratory has extensive experience with both<sup>6,7</sup>. In principle, the design strategy might just as well have been the other way around, Rop → protein G, but Dalal *et al.* reasoned that a helical protein is easier to come by. To achieve their design goal, they changed half the residues in protein G, based on a local secondary structure prediction algorithm<sup>8</sup>, energy minimization, visual modelling, and — impossible to quantify — reliable intuition. The resultant protein was named Janus.

Several other notable challenge-attempts have been reported. Jones *et al.*<sup>9</sup> had partial success in transforming a small, disulphide,  $\beta$ -sheet protein into a helical hairpin. Also, Yuan and Clarke<sup>10</sup> were somewhat successful in changing an all-helical protein to protein G, using an approach similar to that of Dalal *et al.*<sup>2</sup>. Their paper is a 'must read' for its rich discussion of the design issues. Hovering above these overt attempts like a virtual question mark is Alber's work on helical coiled coils<sup>11</sup> in which the modest substitution of a single residue in the heptad repeat unit is sufficient to switch the oligimerization state of the molecule between dimer, trimer and tetramer<sup>12</sup>. More of this in a moment.

Returning to Janus, I have both a question and a quibble. First the quibble. Protein G is 56 residues long; 28 residues were changed in the transformation to Janus. But in addition, the molecule was extended to include a seven residue tail. In Rop, this unstructured tail enhances solubility (inhibiting the transformation of Rop → glop), and presumably it plays the same role in Janus. Calculating backward from Janus to protein G, 28 residues were changed and 7 deleted, a difference of  $(28+7)/63 = 56\%$  between the two molecules, in excess of the 50% threshold. It is argued implicitly that the unstructured tail should not be counted, recalling a riddle attributed to Abraham Lincoln. To paraphrase: how many residues were changed if we ignore the tail? Answer: 35, because ignoring the tail does not eliminate it.

Solubility is not a trivial matter. The reason why a protein is folded, instead of unfolded or aggregated, is not synonymous with the reason why that fold is unique<sup>13</sup>. The information encoded by a protein sequence is spent on all dimensions of the problem, and changes in any one parameter are rightly charged against the total accumulated balance, with its 50% ceiling. Design deficiencies

in both Jones *et al.*<sup>9</sup> and Yuan and Clarke<sup>10</sup> were due, in large part, to problems with maintaining solubility. In Janus, the seven residue tail remains a loose end, in both senses.

So much for the quibble. Next some unfinished business before turning to the question. Why not confer the award on helical coiled-coils<sup>12</sup>, where a 14% change in sequence (one residue per heptad repeat) can transform the oligimerization state? Eaton Lattman's response to this question — only partially tongue-in-cheek — is that Tom Alber didn't ask. A more conventional answer is that a coiled-coil is not a globular protein, and further, a change in oligimerization state is not a change in fold. But such disclaimers are wide open to question. A coiled-coil has a sequence composition, axial ratio, buried interior, and packing density like that of a globular protein. Wherein lies the difference? Again, a Rop dimer has a different fold than cytochrome *b<sub>562</sub>*, another four-helix bundle, so why is a coiled-coil dimer not judged to have a different fold than the corresponding tetramer? Clearly, the Lattman response requires a more searching answer.

At root, the difference between a coiled-coil and a globular protein is that the former lacks peptide chain turns. This difference is not mere phenomenology. It is often argued that turns are passive and therefore neglectable, but this common belief cannot be true. Consider the following thought experiment. An idealized dimer is dissociated into two rigid monomers. In one case, let the monomers be independent (case I), while in another, let them be linked (case L) by a mass-less tether (that is, a turn). Upon dissociation,  $S_{\text{mix}}^{\text{I}} > S_{\text{mix}}^{\text{L}}$ , where  $S_{\text{mix}}$  is the entropy of mixing. Now cut the tether in case L and assume that the work required to do so is precisely offset by the heat of tether breaking. Then, with no expenditure of work, case L is transformed to case I, and the difference,  $\Delta S_{\text{mix}} = S_{\text{mix}}^{\text{I}} - S_{\text{mix}}^{\text{L}}$ , can be ascribed entirely to cratic entropy. Physically, this cratic contribution,  $\Delta S_{\text{mix}}$ , is prepaid in case L by tether formation. Therefore, even a featureless turn matters.

Unlike its Platonic counterpart, a polypeptide tether has mass, stiffness, excluded volume, *etcetera*. In the real case, the prepaid information, corresponding to  $\Delta S_{\text{mix}}$  in the ideal case, will be modulated by the actual sequence, which either favours or opposes turn formation. In a coiled-coil dimer, interactions that local-

ize to the dimer interface are independent, while in a corresponding monomer interface, such interactions are coupled to the connecting tether, which is an obligate repository for such structural information. Unavoidably, the balance between long and short-range forces is shifted in the monomer due to the existence of the tether.

The frequent occurrence of turns is responsible for the globularity of globular proteins<sup>14</sup>. Isodirectional segments ( $\alpha$ -helices or  $\beta$ -strands) are bracketed by turns, which are often inextricably linked to segment termini, for example in helix capping<sup>15,16</sup>. Turns do matter.

Like the Mexican children's game *bale-roc*, where the essence of the game is in the tether\*, the parameters needed to specify protein conformation are reduced when turns are eliminated. Of course, the interface between monomers still needs to be tight enough to thwart aggregation, as it is in a coiled-coil. In a progression that starts in a coiled-coil and finishes in a globular protein, Rop dimer seems to lie somewhere along the way. Indeed, the highly successful design strategy of Dalal *et al.*<sup>2</sup> paid careful heed to the 'a' and 'd' positions in the putative heptad repeat, a knowing backward glance that acknowledges their intellectual debt to the coiled-coil.

Now, finally, the question. Is Janus enough of a globular protein to satisfy the ground rules of the Paracelsus challenge? Coiled-coil characteristics notwithstanding, Rop is just one turn shy of a monomer. In comparison, so is ribonuclease S<sup>17</sup>, a protein that has long been a veritable poster-molecule for model globular behaviour. In the end, lingering doubt clouds the issue either way. And unlike the Supreme Court with pornography (they know it when they see it), no one view predominates here because it is a matter of degree, not kind. This said, I hasten to add that any doubt at the margins must not be allowed to obscure the core level of success that has been achieved by Regan and her colleagues. Janus is an authentic molecular Doppelgänger.

Janus the protein, like Janus the word, is a *contranym*, an antonym of itself, whose contrarian existence poses a conundrum for fold recognition. If chosen from a database of natural proteins, two sequences of length ~60 residues would be certified homologous whenever their aligned sequence similarity exceeds 30%<sup>18</sup>. At 50%, Janus and protein G are surely homologues. But at

41%, so are Janus and Rop. Worse yet, the Janus sequence is significantly closer to protein G than to Rop, *but not the Janus structure*, which resembles Rop, not protein G.

Where does all this leave us? Janus, namesake of the Roman god of new beginnings, has been aptly christened.

*George D. Rose is with the Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland, 21205, USA.*

*email: rose@grserv.med.jhmi.edu*

1. Rose, G. D. & Creamer, T. P. *Proteins: Struct. Funct. Genet.* **19**, 1–3 (1994).
2. Dalal, S., Balasubramanian, S. & Regan, L. *Nature Struct. Biol.* **4**, 548–552 (1997).
3. Aurora, R., Creamer, T. P., Srinivasan, R. & Rose, G. D. *J. Biol. Chem.* **272**, 1413–141 (1997).
4. Gronenborn, A. M. *et al.* *Science* **253**, 657–661 (1991).
5. Banner, D. W., Kokkinidis, M. & Tsernglou, D. *J. Mol. Biol.* **196**, 657–675 (1987).
6. Smith, C. K., Withka, J. M. & Regan, L. *Biochemistry* **33**, 5510–5517 (1994).
7. Munson, M., O'Brien, R., Sturtevant, J. M. & Regan, L. *Protein Science* **3**, 2015–2022 (1994).
8. Garnier, J., Osguthorpe, D. J. & Robson, B. *J. Mol. Biol.* **120**, 97–120 (1978).
9. Jones, D. T. *et al.* *Proteins: Struct. Funct. Genet.* **24**, 502–513 (1996).
10. Yuan, S.-M. & Clarke, N. D. *Proteins: Struct. Funct. Genet.*, submitted.
11. Woolfson, D. N. & Alber, T. *Protein Science* **4**, 1596–1607 (1995).
12. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. *Science* **262**, 1401–1407 (1993).
13. Lattman, E. E. & Rose, G. D. *Proc. Natl. Acad. Sci. USA* **90**, 439–441 (1993).
14. Rose, G. D., Gierasch, L. M. & Smith, J. A. *Adv. Prot. Chem.* **37**, 1–109 (1985).
15. Presta, L. G. & Rose, G. D. *Science* **240**, 1632–1641 (1988).
16. Richardson, J. S. & Richardson, D. C. *Science* **240**, 1648–1652 (1988).
17. Wyckoff, H. W., Tsernglou, D., Hanson, A. W., Knox, J. R., Lee, B. & Richards, F. M. *J. Biol. Chem.* **245**, 305–328 (1970).
18. Sander, C. & Schneider, R. *Proteins: Struct. Funct. Genet.* **9**, 56–68 (1991).

\*A familiar game played with a slender stick and squat ball, attached by a short tether. The stick fits snugly into a slot in the ball. Holding the seated assembly only by the stick, the object of the game is to toss the ball in the air, then catch it with the stick. Considerable skill is required to accomplish this seemingly simple feat.