NATURE|Vol 437|22 September 2005

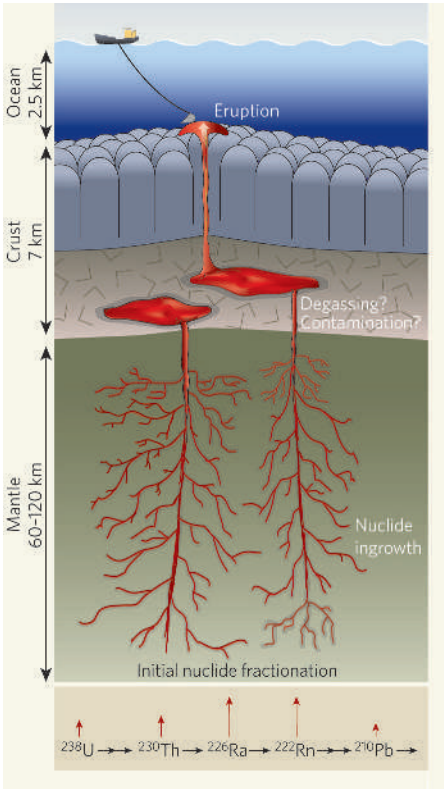

Figure 1 | Melt pathways and possible sites for generation of uranium-series nuclide disequilibrium. Nuclides are fractionated at the onset of melting because of their different affinity for melt relative to the solid. The ^{210}Pb–^{226}Ra disequilibrium in magmas measured by Rubin *et al.*[1] potentially records this event and so the transit time to eruption. Yet continued equilibration between upwelling melt and solid leads to different velocities for the nuclides through the mantle[7], generating further disequilibrium. Generation of disequilibrium by such an 'ingrowth' process is also only effective where the proportion of the melt is small compared to the solid. That is unlikely to be the case in the main melt conduits, but tributaries to the main channels may contribute ingrown nuclides even high in the melting column. Finally, disequilibrium may be caused by contamination or degassing in the crust, but Rubin *et al.* make a good case against this.

The previous speed limit for this process was clocked in 1988 — also by Rubin, who, together with J. D. Macdougall[5], used the ^{226}Ra–^{230}Th pair, which returns to equilibrium in about 8,000 years. The time constraints of this earlier study were thus some two orders of magnitude less stringent than those of the new observations, but at the time they came as a big surprise. In response to the perceived difficulty of moving melt so fast to the surface[6], melting models were developed that relieved some of the need for speed[7]. However, the less glamorous but quite plausible alternative of crustal contamination has also continually raised its head (Fig. 1).

Importantly, the new study[1] not only requires faster melt transport than before but also provides evidence against some of the increasingly sophisticated scenarios of contamination[8]. The effects of contamination on the ^{226}Ra–^{230}Th pair are strikingly different from those on ^{210}Pb–^{226}Ra. Thus, models constructed to explain previously observed ^{226}Ra–^{230}Th excesses by contamination seem unlikely to be able to account for the new ^{210}Pb–^{226}Ra deficits. On the other hand, coupled ^{210}Pb–^{226}Ra deficits and ^{226}Ra–^{230}Th excesses are expected for most melting processes. Rubin *et al.* demonstrate that a simple model can reasonably account for their observations.

Clearly, a more comprehensive exploration of the new data using refined models[9,10] will follow. But now, even more emphatically than before, it seems that you can't keep a good melt down. ■

Tim Elliott is in the Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK.
e-mail: tim.elliott@bristol.ac.uk

1. Rubin, K. H., van der Zander, I., Smith, M. C. & Bergmanis, E. C. *Nature* **437,** 534–538 (2005).
2. Kelemen, P. B., Hirth, G., Shimizu, N., Spiegelman, M. & Dick, H. *Phil. Trans. R. Soc. Lond. A* **355,** 283–318 (1997).
3. Phipps Morgan, J. & Holtzman, B. K. *Geochem. Geophys. Geosyst.* **6,** Q08002; doi:10.1029/2004GC000818 (2005).
4. Condomines, M., Morand, P. & Allègre, C. J. *Earth Planet. Sci. Lett.* **55,** 247–256 (1981).
5. Rubin, K. H. & Macdougall, J. D. *Nature* **335,** 158–161 (1988).
6. Faul, U. *Nature* **410,** 920–923 (2001).
7. Spiegelman, M. & Elliott, T. *Earth Planet. Sci. Lett.* **118,** 1–20 (1993).
8. Saal, A. E. & van Orman, J. A. *Geochem. Geophys. Geosyst.* **5,** Q02008; doi:10.1029/2003GC000620 (2004).
9. Lundstrom, C. *Phys. Earth Planet. Inter.* **121,** 189–204 (2000).
10. Jull, M., Kelemen, P. B. & Sims, K. *Geochim. Cosmochim. Acta* **66,** 4133–4148 (2002).

STRUCTURAL BIOLOGY

Form and function instructions

Jeffery W. Kelly

How much and what kind of information is required to fold a chain of amino acids into a functioning protein? It seems the problem may not be as daunting as once thought — the solution is in the coevolution data.

The linear sequence of amino acids in a protein specifies its final three-dimensional structure and function[1]. But what molecular information is necessary and sufficient to specify protein form and function? In papers on pages 512 and 579 of this issue, Ranganathan and colleagues[2,3] demonstrate that maintaining the conservation pattern in a protein family, along with a surprisingly small subset of coevolving residues, enables the generation of low-homology sequences that fold and function. The studies indicate that the number of crucial interactions in a protein may be smaller than previously thought — a boon for those who want to design novel proteins from scratch to fulfil a specific function.

The authors studied a large family of protein modules, called the WW domains, that mediate protein–protein interactions by binding to sequences that are rich in the amino acid proline[4]. They aligned 120 WW domains from natural proteins, and looked at the distribution of amino acids that occurs at all of the positions along the polypeptide. By comparing each position against the mean distribution from all proteins, they identified those positions that have been conserved throughout evolution and are therefore likely to have some structural or functional significance (Fig. 1). Conservation in these terms means that the amino acid is the same.

They next identified amino-acid positions within the conserved set that seem to have evolved in concert. For example, looking at the 74 sequences with glutamate at position 8 allows comparisons with other positions. Position 16 is a 'conserved' position, but it exhibits a mean distribution of residues in the 74 Glu8 sequences, revealing that position 16 is not coupled to Glu8. However, some conserved positions are coupled to Glu8 — that is, they have coevolved. For instance, position 23 has sequence bias in the 74 Glu8 sequences that is distinct from that at the same position in the 120-sequence alignment.

Position 8 is statistically coupled ($P < 0.05$) to just six other sites in the sequence, and a matrix of evolutionary couplings for several positions shows a pattern where a few positions are mutually conserved and most positions interact only weakly, if at all. This result was surprising, and suggested that a computer program might be able to design novel WW domain sequences using conservation and coupling information only, without any information about three-dimensional structure.

To test this idea, Ranganathan and colleagues[2,3] developed two programs: 'algorithm 1' generates artificial WW domain sequences that preserve amino-acid conservation, but eliminates all statistical coupling between sites; and 'algorithm 2' generates sequences that retain both statistical conservation and coupling. They produced four libraries of synthetic genes that encode the WW domain. The first library encodes 42 natural WW domains as a control, the second was produced by algorithm 1, the third set was built by algorithm 2,

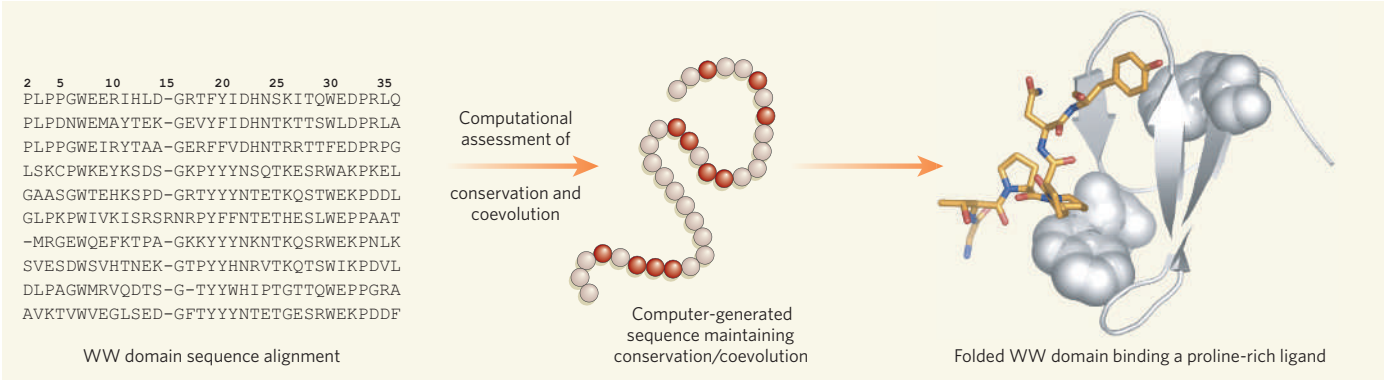

Figure 1 | The grand plan. Ranganathan and colleagues[2,3] aligned the sequences of WW protein domains and computationally identified those residues showing conservation and coevolution. Computer programs that preserved this information were used to generate libraries of synthetic sequences with modest homology, which were tested experimentally for their ability to fold and function. Conserved and coevolving amino-acid residues are shown as red spheres.

and the fourth set contained 19 random sequences as a negative control.

The authors then used the libraries to express the genes in terms of amino-acid sequence and tested whether they could fold and function like the natural domains. Of the natural sequences, 67% folded in the experimental conditions used. Notably, 25% of the sequences produced from algorithm 2 (conservation + coevolution) could fold, but none of the algorithm-1 sequences (just conservation) could. The sequences generated by algorithms 1 and 2 had analogous extents of identity with natural WW domains, including positions within the hydrophobic core. Moreover, the artificial WW domains function like their natural counterparts: 60% of the algorithm-2 sequences exhibited class-specific binding to proline-rich sequences, with affinities analogous to those seen in many of the original 120 domains.

So it seems that both conservation and evolutionary-coupling information is necessary and sufficient to produce sequences with modest similarity to native sequences that can fold and function. Despite the myriad of local and more-distant interactions revealed by three-dimensional protein structures, these studies imply that there are only a few crucial 'coupled' interactions in a sea of weaker interactions. In other words, there may be a few highly ordered, energetically important molecular interactions that are mixed with many more-fluid, less-important interactions. It is not surprising that these would be difficult to distinguish by simply inspecting averaged structures. It is interesting that many of the conserved and coevolving residues in WW domains are important for fold stability or function[5,6]. In contrast, the residues that have been shown to control the kinetics of folding are not conserved or coupled, suggesting that folding rate is not optimized by evolution[5,6].

Ranganathan and colleagues' algorithm will undoubtedly be tested by the protein-design community in cases for which genetic information is available. The statistical-coupling matrix will allow hypotheses about the evolutionary constraints on a protein to be tested experimentally. The energetics of folding and function are expected to be manifest as statistical couplings, and it remains to be seen whether these and other features are distinct, overlapping or inseparable features of the total sequence information. It will be interesting to find out whether key conformational changes associated with enzyme and receptor function are preserved in the evolutionary record[7,8]. Most importantly, this statistical approach, which does not rely on three-dimensional structural information, has the potential to identify crucial residues that perform feats in proteins that we do not currently appreciate. ■

Jeffery W. Kelly is in the Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, USA.
e-mail: jkelly@scripps.edu

1. Anfinsen, C. B. *Science* **181,** 223–230 (1973).
2. Socolich, M. *et al. Nature* **437,** 512–518 (2005).
3. Russ, W. P., Lowery, D. M., Mishra, P., Yaffe, M. B. & Ranganathan, R. *Nature* **437,** 579–583 (2005).
4. Sudol, M., Chen, H. I., Bougeret, C., Einbond, A. & Bork, P. *FEBS Lett.* **369,** 67–71 (1995).
5. Jager, M., Nguyen, H., Crane, J. C., Kelly, J. W. & Gruebele, M. *J. Mol. Biol.* **311,** 373–393 (2001).
6. Deechongkit, S. *et al. Nature* **430,** 101–105 (2004).
7. Shulman, A. I., Larson, C., Mangelsdorf, D. J. & Ranganathan, R. *Cell* **116,** 417–429 (2004).
8. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. *Science* **295,** 1520–1523 (2002).

SYNTHETIC CHEMISTRY

Recipes for excess

John Hartwig

The selective production of a particular mirror-image form of a molecule is immensely important to organic synthesis. But techniques to find the right catalysts have traditionally been protracted and fiddly. Help is at hand.

Chiral molecules are molecules that come in two non-superimposable mirror-image forms, known as enantiomers. Synthesizing one enantiomer of a chiral molecule in preference to the other is difficult but crucial: among other things, single-enantiomer drugs account for some 40% of worldwide drug sales, worth more than US$100 billion[1]. In recognition of this, the 2001 Nobel Prize in Chemistry[2] was awarded for the development of chiral catalysts for 'enantioselective' synthesis. A notable advance since then has been the recognition that the often neglected, more symmetrical 'achiral' components of a chiral catalyst can be used to improve enantioselectivity further. Recent work[3,4] that focuses on transformations involving rhodium phosphite and phosphoramidite catalysts is illustrative of this approach.

Living organisms contain countless chiral biomolecules: proteins, nucleic acids and carbohydrates, not to mention enzyme cofactors, vitamins and other trace substances. On their own, the enantiomers of these molecules — and the enantiomers of small, chiral organic compounds — have the same properties. The distinctive properties of the two enantiomers arise only when they interact with other chiral molecules. This is best explained by the analogy of a handshake. When two people shake hands, they both extend their right hand: a handshake with two right hands fits better than one with a right and a left. Similarly, a 'right-handed' enantiomer of a drug might fit a right-handed 'glove' (binding pocket) of a protein better than the 'left-handed' version of the same drug. The activity of a small-molecule drug (with a molecular mass typically less than 500 daltons) is therefore critically affected by its handedness — its chirality.