sequence, single nucleotide polymorphism (SNP), haplotype, linkage and northern blot analysis, these groups identified missense mutations R1441G, R1441C, Y1699C, I1122V and I2020T segregating with PD with high penetration. Furthermore, these mutations were absent in over 1000 control individuals, indicating that these are pathogenic mutations and not mere polymorphisms.

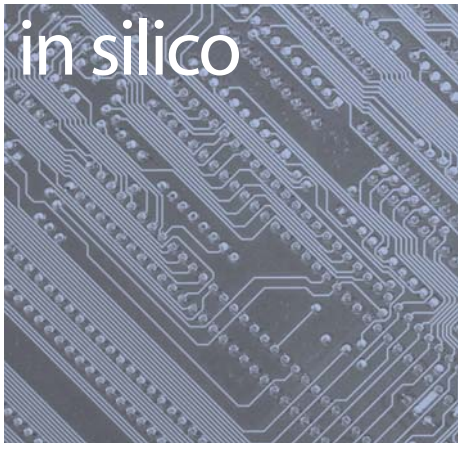
The gene product of *PARK8/LRRK2* is dardarin/LRRK2, predicted to consist of 2527 amino acids encompassing 12 leucine-rich repeats, a tyrosine kinase-like domain, a RAS/small GTPase superfamily domain and a WD40 domain. It is of particular interest that the mutations R1441G and R1441C are within the GTPase domain, given that Ras-like small GTPases can act as molecular switches regulating gene expression, vesicle trafficking, nucleocytoplasmic transport, mitogenic signaling as well as microtubule organization. Although speculative at this point, it is possible that at least some of these cellular functions are affected by the mutation of the conserved arginine residue. Furthermore, the mutation I2020T is within the kinase domain. Another attractive hypothesis proposes that LRRK2 is responsible for the phosphorylation of α-synuclein, thereby playing a key role in the deposit of this protein in dying neurons.

By delineating the role of LRRK2 in the pathogenesis of PD, we will be closer to defining the underlying mechanism for this multifactorial disease. Insights into the disease process would enable us to select druggable targets based on rational molecular approaches. The identification of disease-segregating mutations in *PARK8/LRRK2* has indeed opened up new doors for the search of therapeutics that would prevent or ameliorate PD.

6 Paisán-Ruíz, C. *et al.* (2004) Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* 44, 595–600

7 Zimprich, A. *et al.* (2004) Mutations in LRRK2 cause autosomal-dominant Parkinsonism with pleomorphic pathology. *Neuron* 44, 601–607

Herman H. Cheung
herman@mgcheo.med.uottawa.ca



Beware generalizations!

In 2002, researchers at GSK in Pennsylvania published the results of an analysis of rat oral bioavailability (%F) data for over 1100 drug candidates. One of the conclusions from their work was that 80% of compounds meeting two simple criteria showed a %F value of greater than or equal to 20%. The two criteria were that a compound should possess 10 or fewer rotatable bonds and have a polar surface area of 140 Å^2 or below. In the wake of Lipinski's 'rule-of-five', such simple rules-of-thumb have been eagerly seized on because they appear to offer a rapid means of prioritizing compounds at an early stage of drug discovery.

However, in a recent publication [1], a group of scientists from the former Pharmacia site in Kalamazoo has performed a similar analysis on rat %F data for 434 Pharmacia compounds and arrived at some interesting conclusions. First, perhaps obviously to computational scientists, the precise values that are obtained for the number of rotatable bonds in a molecule and its polar surface area depend upon the software that is used to compute these quantities. So, unless it is certain that the software used is the same as in the work used to derive the 'rule', then there is always the danger that a compound will be misclassified.

Second, even when the Pharmacia scientists tried to mimic the GSK calculation procedure as closely as possible, they observed that the percentage of their compounds with %F >20 that satisfied the two criteria was only 70%, compared with the 80% figure reported by the GSK group. Furthermore, the Pharmacia group noticed the results varied depending on the therapeutic target against which the compounds were directed.

The overall conclusions from the work were twofold: (1) it is likely to be extremely difficult to obtain a simple and general rule to predict a complex physiological endpoint such as oral bioavailability; (2) that great caution should be exercised when prospectively applying conclusions obtained from one set of data using one particular computational protocol to other data sets using different protocols.

1 Lu, J.J. *et al.* (2004) Influence of molecular flexibility and polar surface area metrics on oral bioavailability in the rat. *J. Med. Chem.* 47, 6104–6107

David Clark
David.Clark@ArgentaDiscovery.com

Unravelling the ORFan mystery

With whole-genome sequencing projects progressing fast, studies are well underway to elucidate evolutionary relationships among different organisms on a molecular basis. However, ~20–30% of newly identified open reading frames (ORFs) from a sequenced genome lack any detectable sequence similarity to known ORFs in other species. The number of these orphan ORFs – so called ORFans – can rise up to 60% as in the case of the malaria parasite *Plasmodium falciparum*, making functional and structural assignment difficult using common bioinformatics approaches. ORFans become increasingly abundant in sequence databases with >30,000 representatives to date. Much has been speculated about their origins and roles proposed ranging from 'sequencing errors' to 'non-expressed pseudogenes' to 'unique proteins with new function/3D structure' as few ORFans have been studied experimentally.

Siew and Fischer are challenging this ORFan mystery, describing the impact of structural biology on ORFan research by analyzing recent PDB entries for sequence homology to known proteins [2]. They identified 11% of these as ORFans from various Kingdoms (Archea, Bacteria, Eukarya) and viruses. This suggests that: (1) most ORFans are likely to be expressed into functional proteins; and (2) ORFans are already commonly studied experimentally and more frequently than previously expected. In fact, ~75% of these identified ORFans are already functionally characterized, many assuming roles in transcription and/or translation.

In terms of structure the researchers found that most ORFans have a known fold. The team questioned if fold-recognition methods would have been able to predict those structures with a common fold and found that 30% were correctly predicted when compared with the experimental structure. These findings highlight once more that structure is more conserved than sequence in evolution and the key to function prediction and unfolding evolutionary origins.

The snapshot survey on structural ORFan research undertaken by these authors demonstrates the impact of structural biology on assigning functions and unravelling evolutionary relationships and origins for putative genes discovered through genome-sequencing projects without significant sequence relationship to known genes. Large-scale structural studies are needed to characterize the remaining thousands of ORFans and will show if ORFans are promising source of novel folds and attractive targets for further studies.

2 Siew, N. and Fischer, D. (2004) Structural biology sheds light on the puzzle of genomic ORFans. *J. Mol. Biol.* 342, 369–373

Monika Rella
bmbmre@bmb.leeds.ac.uk

1359-6446/04/$ – see front matter PII: S1359-6446(04)03348-3