

## OPINION

# How many drug targets are there?

John P. Overington, Bissan Al-Lazikani and Andrew L. Hopkins

**Abstract** | For the past decade, the number of molecular targets for approved drugs has been debated. Here, we reconcile apparently contradictory previous reports into a comprehensive survey, and propose a consensus number of current drug targets for all classes of approved therapeutic drugs. One striking feature is the relatively constant historical rate of target innovation (the rate at which drugs against new targets are launched); however, the rate of developing drugs against new families is significantly lower. The recent approval of drugs that target protein kinases highlights two additional trends: an emerging realization of the importance of polypharmacology, and also the power of a gene-family-led approach in generating novel and important therapies.

Surprisingly, for an industry that spends in excess of US\$50 billion on research and development each year, there is a lack of knowledge of the set of molecular targets that the modern pharmacopoeia acts on. If we are to develop predictive methods to identify potential new drug targets, it is essential that we establish with confidence the number, characteristics and biological diversity of targets of approved drugs.

Drews and Reiser were the first to systematically pose and answer this question, identifying 483 drug targets<sup>1,2</sup>. Later, an analysis by Hopkins and Groom challenged this figure and suggested that ‘rule-of-five’-compliant<sup>3</sup> drugs acted primarily through only 120 underlying molecular targets<sup>4</sup>. Subsequently, in 2003, Golden proposed that all then-approved drugs acted through 273 proteins<sup>5,6</sup>. By contrast, Wishart *et al.*<sup>7</sup> reported 14,000 targets for all approved and experimental drugs, although they revise this number to 6,000 targets on the DrugBank database website. In 2006, Imming *et al.* catalogued 218 molecular targets for approved drug substances<sup>8</sup>, whereas Zheng *et al.* disclose 268 ‘successful’ targets in the current version of the Therapeutic Targets Database<sup>9,10</sup>. Here, we propose a consensus number of 324 drug targets for all classes of approved therapeutic drugs (TABLE 1), reconciling earlier reports into a current and comprehensive survey. The details of this target list, including the names

of the targets, target class, cellular location and United States Adopted Names (USAN) or launch dates, are illustrated in the accompanying poster on the molecular pharmacopoeia (see Further information online).

In order to produce this consensus on the number of molecular targets, a comprehensive analysis of the US FDA’s Orange Book (for primarily small-molecule drugs) and the Center for Biologics Evaluation and Research (CBER) website (for biological drugs) was performed. This analysis identifies in excess of 21,000 drug products; however, when duplicate active ingredients, salt forms, supplements, vitamins, imaging agents, and so on are removed, this number is reduced to only 1,357 unique drugs, of which 1,204 are ‘small-molecule drugs’ and 166 are ‘biological’ drugs.

Of the 1,204 small-molecule drugs, 803 can be administered orally, 421 can be dosed parenterally and 275 can be used as topical agents (for this analysis, buccal, rectal, inhalational and other such routes of administration are considered as topical agents). A further complicating factor in these analyses is that some drugs are dosed as prodrugs. Interestingly, a significant number — at least 192 (16%) — of the small-molecule drugs are prodrugs. The rule-of-five<sup>3</sup> is routinely used to assess the likelihood of oral absorption for drugs. In our analysis, 885 of all small-molecule drugs pass the rule-of-five test; of these, 619 (70%) are actually dosed orally, whereas 159 (20%) of orally dosed drugs fail at least one of the rule-of-five parameters.

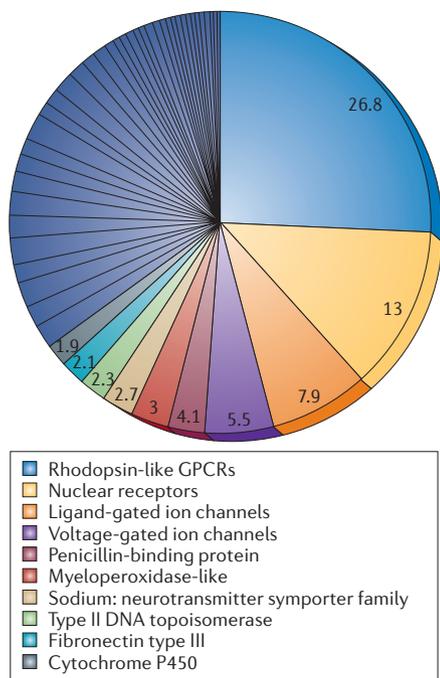
### Assignment of efficacy targets

In order to identify the efficacy targets — the molecular targets through which the drug mediates its approved therapeutic activities — we conducted a comprehensive analysis of the literature for each drug. The criteria for assignment was strict in that strong evidence of cell-based and/or *in vivo* evidence linking the target (and specific target sub-type) to the effect of the drug must exist alongside binding data. We were able to assign protein molecular targets believed to be responsible for the efficacy of a drug to 1,065 of the unique drugs. Where possible, we also recorded the particular domain in the protein that binds the drug, as well as the known binding sites/residues.

The literature is often complex in terms of the information provided about efficacy targets. For cases in which a specific target

Table 1 | **Molecular targets of FDA-approved drugs**

Class of drug target	Species	Number of molecular targets
Targets of approved drugs	Pathogen and human	324
Human genome targets of approved drugs	Human	266
Targets of approved small-molecule drugs	Pathogen and human	248
Targets of approved small-molecule drugs	Human	207
Targets of approved oral small-molecule drugs	Pathogen and human	227
Targets of approved oral small-molecule drugs	Human	186
Targets of approved therapeutic antibodies	Human	15
Targets of approved biologicals	Pathogen and human	76



**Figure 1 | Gene-family distribution of current drugs per drug substance.** The family share as a percentage of all FDA-approved drugs is displayed for the top ten families. Beyond the ten most commonly drugged families, there are a further 120 domain families or singletons for which only a few drugs have been successfully launched. Data based on 1,357 dosed components from >20,000 approved products, FDA, December 2005. GPCR, G-protein-coupled receptor.

protein is believed to be the sole or major route through which a drug achieves its efficacy, we assign the drug against this single target; for example, the histamine H<sub>1</sub> receptor is believed to be the major mechanistic target for cetirizine and hydroxyzine, and acebutolol acts through the β<sub>2</sub> adrenoceptor, although all these drugs show binding to other G-protein-coupled receptors (GPCRs) in *in vitro* assays. In other cases, the drug acts through a number of target subtypes: for example, carvedilol acts through blocking a number of α- and β-adrenoceptors. Finally, a drug can act through multiple distinct mechanisms, and therefore unrelated targets. For example, ritonavir is an HIV protease inhibitor; however, it is usually given in combination with other HIV protease inhibitors because it inhibits the cytochrome P450 3A4 (CYP3A4)-mediated metabolism of other HIV protease inhibitors such as lopinavir. In such cases, both HIV1 protease and human CYP3A4 are regarded as the molecular targets.

There is a relatively small, but clinically significant, class of drugs that bind to either ribosomes or DNA, or that have no distinct

or an unknown mode of action. The literature changes frequently in terms of the knowledge available about drug indications and mechanisms of action, and so this information needs to be reviewed regularly.

**Efficacy targets of current drugs**

On the basis of existing knowledge, we were able to determine that all current drugs with a known mode-of-action act through 324 distinct molecular drug targets. Of these, 266 are human-genome-derived proteins, and the remainder are bacterial, viral, fungal or other pathogenic organism targets. Small-molecule drugs modulate 248 proteins, of which 207 are targets encoded by the human genome (TABLE 1). Oral small-molecule drugs target 227 molecular targets, of which 186 are human targets.

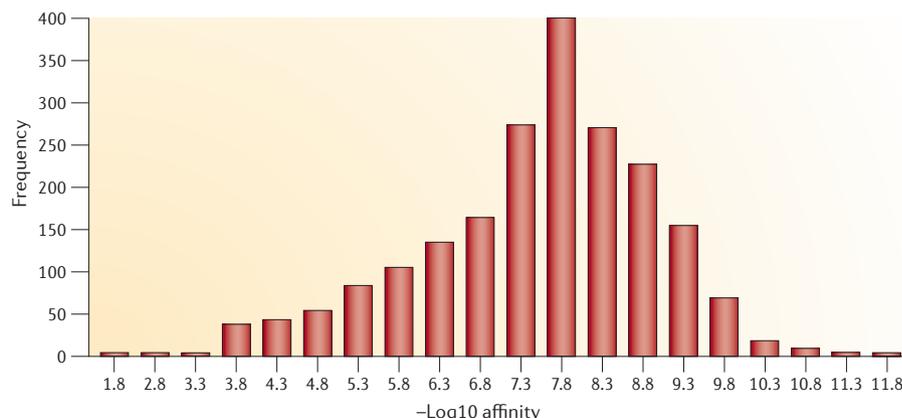
A complicating feature of any such analysis is that many drugs have complex and relatively poorly understood pharmacology, and some targets are actually complex multimeric proteins with variable subunit compositions and so on. If one makes the assumption that proteins related down to 50% identity show related pharmacology, then this list of 324 targets expands to 604 genes for the human genome (comparison carried out against ENSEMBL genome June 2006 release containing 29,679 genes). Extending the analysis to include all close homologues (35% identity or closer) increases the number to 1,048 genes (3.5% of the genome). This line of reasoning lead to the initial estimate of the size of the druggable genome<sup>4</sup>. Understanding the real pharmacological footprint of current drugs offers many opportunities for both developing new, optimized agents with different selectivity profiles, and also more efficient lead discovery and optimization strategies.

Current biological drugs target 76 proteins, with currently marketed monoclonal

antibody therapeutics acting on 15 distinct human targets. So far, only nine targets are modulated by both small-molecule and biological drugs, with the differing agent types usually targeting different domains or binding sites. This relatively small number of jointly modulated targets is driven by both technical and commercial considerations. For example, the biological drugs cetuximab and panitumumab target the extracellular domain of the receptor tyrosine kinase EGFR (ERBB1), whereas the small-molecule drugs gefitinib and erlotinib target the adenine portion of the ATP-binding site of the cytosolic catalytic kinase domain within the same receptor.

**Drug polypharmacology**

It was clear from both our curation of drug targets from the literature and also data-mining of known affinity values of drugs for targets (as abstracted in a large database of medicinal chemistry literature<sup>17</sup>) that many drugs show clinically relevant polypharmacology (that is, they are ‘dirty drugs’<sup>18</sup>). Quite expectedly, closely related members of the gene family will show significant drug promiscuity, and, as a result of the generally similar function of these proteins, give rise to complex composite clinical pharmacology. The point of genuine multitarget effects of drugs is well illustrated by several recently launched protein kinase inhibitors. Imatinib, originally developed as a highly selective inhibitor of c-ABL<sup>11</sup> (and which target association led to its first approval for chronic myeloid leukaemia), has subsequently been discovered to be have significant activity against several other clinically relevant kinases, such as c-KIT<sup>12–14</sup>, leading to expansion of the clinical utility of this important therapeutic. Sorafenib has been recently launched as an explicit multikinase inhibitor, affecting both tumour proliferation and tumour angiogenesis pathways, and acting



**Figure 2 | Frequency distribution for small-molecule drug potencies.**

through at least RAF-kinase, vascular endothelial growth factor 2 (VEGFR2), VEGFR3, platelet-derived growth factor receptor- $\beta$  (PDGFR  $\beta$ ), KIT and FLT3<sup>15</sup>. Finally, sunitinib is a further recently approved protein kinase inhibitor with a similar target footprint to sorafenib<sup>12</sup>.

At a reductionist level, this polypharmacology can be rationalized by similarities of key pharmacophores in the binding sites of proteins from different families<sup>16</sup>, and, when comparing proteins from within the same protein family, polypharmacology can be rationalized by similarities or differences in the physicochemistry of the equivalent residues lining the binding site. The recent work on large-scale mapping of polypharmacology interactions reveals the extent of promiscuity of drugs and leads across the proteome<sup>17</sup>. Placing such aspects of polypharmacology on a firm practical and theoretical footing is a major challenge for the development of future safe and clinically differentiated therapies<sup>17,18</sup>.

### Patterns in drug targets

In order to identify the familial relationships between all drug targets, we analysed the presence of domains, using the SCOP<sup>19</sup> and PFAM<sup>20</sup> databases. Approximately 130 'privileged druggable domains' cover all current

### Glossary

#### Cytochrome P450 3A4

This enzyme is arguably the most important enzyme for drug metabolism; it metabolizes more than 50% of marketed drugs, and is frequently involved in drug–drug interactions.

#### New molecular entity

A drug that contains an active ingredient that has not been previously approved by the US FDA.

#### Pharmacophore

The ensemble of steric and electronic features that is necessary to ensure optimal interactions with a specific biological target structure and to trigger (or to block) its biological response.

#### Polypharmacology

Here we use polypharmacology to mean the binding of a drug to multiple target proteins, with clinical effects being mediated through the modulation of the set of protein targets.

#### Privileged druggable domains

A functional domain of a protein for which a significant fraction of family members have been successfully targeted by drugs. Rhodopsin-like GPCRs, certain ion-channel domains and nuclear receptor ligand-binding domains are clear historical examples of druggable domains.

#### Prodrug

A drug that requires conversion to a more active pharmacological form following dosing. This conversion is often performed by endogenous enzymes. Prodrugs are generally used to overcome problems with stability, toxicity or often limited oral bioavailability of the pharmacologically active form.

#### Rule-of-five

Poor absorption or permeation of a compound is more likely when there are >5 hydrogen bond donors, the molecular mass is >500, cLogP is >5, and the sum of nitrogen and oxygen atoms in a molecule is greater than 10. Many drugs, however, are exceptions to the rule-of-five, and often these are substrates for biological transporters.

drug targets. This number is in stark contrast to the projected number of protein families and folds (10,000 folds<sup>21</sup> and more than 16,000 families<sup>22</sup>).

Analysis of the gene-family distribution of targets by drug substance for both small-molecule and biological drugs (FIG. 1) reveals that more than 50% of drugs target only four key gene families: class I GPCRs, nuclear receptors, ligand-gated ion channels and

voltage-gated ion channels. The targets with the most number of drugs approved are the glucocorticoid receptor and the histamine H<sub>1</sub> receptor (see accompanying poster; also available online). Protein kinases are relatively over-represented considering the small number of launched drugs against these targets: there are currently five approved drugs against the protein catalytic domain, acting on 18 protein kinases.

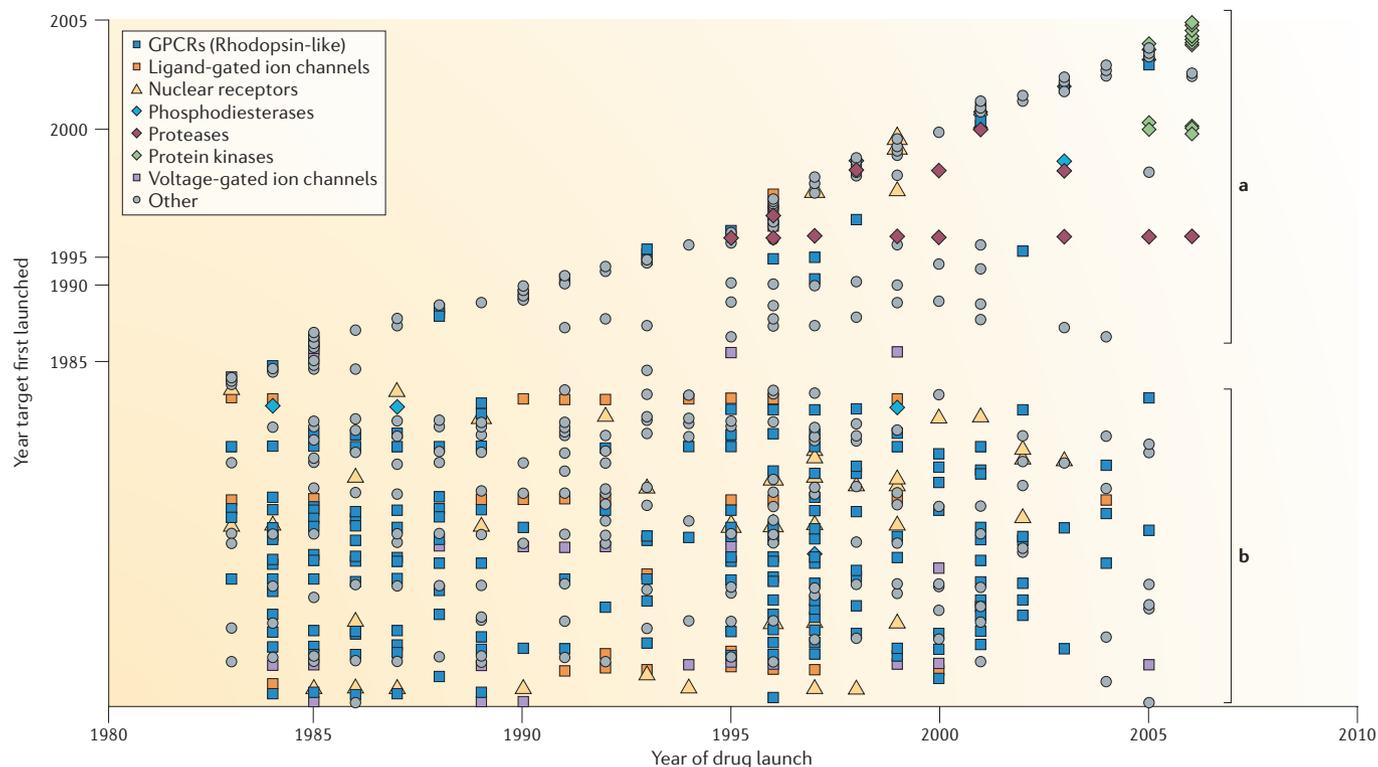


Figure 3 | **Rate of target innovation.** The y-axis represents the year of first drug launch against each target, and the x-axis is the year of each subsequent drug release, with the plot ordered so that more recently 'drugged'

targets are shown at a higher y ordinate. Region **a** reflects periods of high target innovation (after 1982) while region **b** is predominantly the re-use of established mechanisms. The rate of new protein families per year is 1.9.

A common property of most current drug targets is their cell-surface location: 60% of drug targets are located at the cell surface, compared with only ~22% of all proteins in the human genome. Despite this strong pattern of membrane association, there is a high degree of structural characterization of drug targets, with 105 drug targets themselves and over 92% (300) of drug targets being similar to known proteins in the Protein Data Bank (PDB)<sup>23</sup>. Although it is still highly debatable whether useful design and docking can be performed in all cases, this high structural coverage clearly illustrates the progress structural biology has been making in its application to drug discovery.

The Online Mendelian Inheritance of Man (OMIM) database<sup>24</sup> provides associations between genetic loci and diseases. A smaller number of disease associations are linked directly to a specific gene. In an analysis of these data, 1,620 distinct human protein sequences are linked directly to a genetic disease. Of these, 105 are drug targets, corresponding to 47% of human drug targets that are directly associated with a disease. In an analysis against an earlier version of the human genome, only 7% (1,620 out of 22,218) genes have explicit disease associations.

The availability of this carefully curated database allowed us to perform further analyses that address other features of the properties of marketed small-molecule drugs of broad relevance. One example is of the potency distribution of marketed drugs. We identified binding-affinity-related endpoints for all drug-efficacy target pairs identified in the above analyses. The extracted data included all  $IC_{50}$ ,  $EC_{50}$ ,  $ED_{50}$ ,  $K_p$ ,  $K_d$  and  $pA_2$  data; data were transformed as appropriate onto a log affinity scale and are detailed in FIG. 2. As can be seen, the median affinity for current small-molecule drugs is around 20 nM.

One striking feature of the analysis is how infrequent it is for new protein domains to join the exclusive drugged domain set. Using the same dataset analysed by a recent National Institute for Health Care Management Research and Educational Foundation (NIHCM Foundation) study examining drug innovation<sup>25</sup>, of 361 new molecular entities (NMEs) approved by the FDA between 1989 and 2000, 76% targeted a precedented drugged domain and only 6% targeted a previously undrugged domain; the remainder have either unknown targets (4%) or are believed not to have distinct molecular targets underlying their action (17%). Analysis of the rate of target innovation is shown in FIG. 3. As can be seen, the rate

is surprisingly constant over the past 20 years, with an average rate of first-against-target drugs of 5.3 new 'drugged' targets per year, although over individual years the rate is quite variable. The recent set of multikinase-targeted drugs are clearly a significant modern pattern, following the 2001 launch of imatinib. The first approved indications for drugs acting on the new drug targets are commonly orphan diseases, suggesting that innovation primarily occurs at the edges of the market (as is typical behaviour of innovation in other industries<sup>26</sup>), whereas innovations in the major disease indications can occur as rarely as once a decade, owing to the challenge of showing superior safety and efficacy profiles over existing therapies with very large and costly outcome-based clinical trials.

### Concluding thoughts

It is clear from this work, and the recent work of others, that despite the huge efforts and resources aligned to drug discovery as a commercial process, there is relatively little publicly accessible data addressing the fundamental question of how current drugs work. We hope this perspective helps address some of the issues surrounding our current knowledge of the complexities of drug action, and also provides a useful assembly of data that could only have been achieved through careful consolidation of literature, and other disparate data sources. Given this collection of historical drug targets, it is now possible to start to address the development of computer-based approaches to rationally score and assess the likely success of future drug targets. It is tempting to see advances in technology as transforming the rate of innovation in discovery, but there has been little evidence of technological developments to date improving our ability to tackle new target classes with increasing speed and success, and it would seem that small, incremental changes and hard-won gains in fundamental biology and chemistry, and clinical science, is still state of the art. As James Black is famously quoted as saying: "the most fruitful basis for the discovery of a new drug is to start with an old drug"<sup>27</sup>.

John P. Overington and Bissan Al-Lazikani  
are at Inpharmatica Ltd., 1 New Oxford Street,  
London, WC1A 1NU, UK.

Andrew Hopkins is at Pfizer Global Research  
and Development, Sandwich, Kent, CT13 9NJ, UK.

Correspondence to J.P.O.  
e-mail: j.overington@inpharmatica.co.uk

doi:10.1038/nrd2199

1. Drews, J. Genomic sciences and the medicine of tomorrow. *Nature Biotechnol.* **14**, 1516–1518 (1996).

2. Drews, J. & Ryser, S. Classic drug targets. *Nature Biotechnol.* **15**, 1318–1319 (1997).
3. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* **23**, 3–25 (1997).
4. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nature Rev. Drug Discov.* **1**, 727–730 (2002).
5. Golden, J. B. Prioritizing the human genome: knowledge management for drug discovery. *Curr. Opin. Drug Discov. Dev.* **6**, 310–316 (2003).
6. Golden, J. Towards a tractable genome: knowledge management in drug discovery. *Curr. Drug Discov.* **17**–20 (2003).
7. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* **43**, D668–D672 (2006).
8. Imming, P., Sinning, C. & Meyer, A. Drugs, their targets and the nature and number of drug targets. *Nature Rev. Drug Discov.* **5**, 821–834 (2006).
9. Zheng, C., Han, L., Yap, C. W., Xie, B. & Chen, Y. Progress and problems in the exploration of therapeutic targets. *Drug Discov. Today* **11**, 412–420 (2006).
10. Zheng, C. J. *et al.* Therapeutic targets: progress of their explorations and investigation of their characteristics. *Pharma. Rev.* **58**, 259–279 (2006).
11. Druker, B. J. *et al.* Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nature Med.* **2**, 561–566 (1996).
12. Fabian, M. A. *et al.* A small molecule-kinase interaction map for clinical kinase inhibitors. *Nature Biotechnol.* **23**, 329–336 (2005).
13. Atwell, S. *et al.* A novel mode of Gleevec binding is revealed by the structure of spleen tyrosine kinase. *J. Biol. Chem.* **279**, 55827–55832 (2004).
14. Heinrich, M. C. *et al.* Inhibition of c-kit receptor tyrosine kinase activity by STI 571, a selective tyrosine kinase inhibitor. *Blood* **96**, 925–932 (2000).
15. Ahmad, T. & Eisen, T. Kinase inhibition with BAY 43–9006 in renal cell carcinoma. *Clin. Cancer Res.* **10**, 6388S–6392S (2004).
16. Weber, A. *et al.* Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: New pharmacological opportunities due to related binding site recognition. *J. Med. Chem.* **47**, 550–557 (2004).
17. Paolini, G. V., Shapland, R. H. B., van Hoorn, W. P., Mason, J. S. & Hopkins, A. L. Global mapping of pharmacological space. *Nature Biotechnol.* **24**, 805–815 (2006).
18. Hopkins, A. L., Mason, J. S. & Overington, J. P. Can we rationally design promiscuous drugs?. *Curr. Opin. Struct. Biol.* **16**, 127–136 (2006).
19. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **274**, 536–540 (1995).
20. Bateman, A. *et al.* The Pfam Protein Families Database. *Nucleic Acids Res.* **32**, D138–D141 (2004).
21. Koonin, E. V., Wolf, Y. I. & Karev, G. P. The structure of the protein universe and genome evolution. *Nature* **420**, 218–223 (2002).
22. Vitkup, D., Melamud, E., Moul, J. & Sander, C. Completeness in structural genomics. *Nature Struct. Biol.* **8**, 559–566 (2001).
23. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acid Res.* **28**, 235–242 (2000).
24. McKusick, V. A. in *Mendelian Inheritance in Man* 12th Edn (John Hopkins University Press, Baltimore, 1998).
25. NIHCM 24 pp (National Institute for Health Care Management Research and Educational Foundation, 2002).
26. Christensen, C. M. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail* (Harvard Business School Press, Cambridge, 1997).
27. Raju, T. N. The Nobel chronicles. *Lancet* **355**, 1022 (2000).

### Acknowledgements

We thank I. Carruthers, R. Cox, S. Rehman and J. Stevenson for assistance with data curation and analysis.

### Competing interests statement

The authors declare **competing financial interests**: see Web version for details.

### FURTHER INFORMATION

Poster on the molecular pharmacopoeia:  
<http://www.nature.com/focus/molecularpharmacopoeia>  
Access to this links box is available online.