DRUG DISCOVERY TODAY TARGETS

Predicting protein druggability

Philip J. Hajduk, Jeffrey R. Huth and Christin Tse

The ability to predict whether a particular protein can bind with high affinity and specificity to small, drug-like compounds based solely on its 3D structure has been a longstanding goal of structural biologists and computational scientists. The promise is that an accurate prediction of protein druggability can capitalize on the huge investments already made in structural genomics initiatives by identifying highly druggable proteins and using this information in target identification and validation campaigns. Here we discuss the potential utility of tools that characterize protein targets and describe strategies for the optimal integration of protein druggability data with bioinformatic approaches to target selection.

Recent technological advances in NMR spectroscopy and X-ray crystallography have dramatically increased our ability to produce high-resolution 3D structures of proteins either alone or in complex with natural and non-natural ligands. In fact, a survey of the Protein Data Bank (PDB) reveals that there are more than 7,000 entries for unique eukaryotic and prokaryotic proteins (http://dunbrack.fccc.edu/PISCES.php). Although this number is impressive, it still represents only a fraction of the number of proteins that are predicted to be in the proteome, which ranges from ~6000 for prokaryotic organisms to ~30,000 for humans [1]. To close this gap, several structural genomics initiatives have been launched, with the aim of providing structural data on a genomic scale and delineating the total repertoire of protein folds [2–4] (see also http://www.rcsb.org/pdb/strucgen.html). Although many facets of high throughput cloning, expression, purification, and structure determination have yet to be fully implemented and refined, these initiatives are already producing hundreds of structures each year and providing essential data for understanding the function of these gene products in cell biology and disease progression. The availability of structural data, especially of proteins complexed to small molecule ligands, has enabled numerous analyses that attempt to understand and predict the forces that govern molecular recognition and ligand binding [5]. Understanding these principles is important not only for biological insight, but also in the search for drugs that can bind to these proteins with high affinity and specificity to alter their function *in vivo*. To truly exploit the data produced by structural genomics approaches, it is necessary to rapidly and quantitatively assess large numbers of proteins for their potential to be targeted with small molecule therapeutics.

Target validation in the post-genomics era

Successful drug development requires a disease target that plays a vital role in the causation and/or progression of the disease phenotype and that can be modulated with a drug molecule. In other words, therapeutically relevant targets are both 'disease-modifying'

Philip J. Hajduk*
Jeffrey R. Huth
Christin Tse
Pharmaceutical Discovery Division GPRD,
Abbott Laboratories, R46Y, AP-10,
100 Abbott Park Road,
Abbott Park,
IL 60064-3500,
USA
*e-mail: philip.hajduk@abbott.com

1359-6446/05/$ – see front matter PII: S1359-6446(05)03624-X

GLOSSARY

Druggable binding site
A pocket on a protein surface that can bind with high affinity and specificity to small, drug-like molecules.

Nondruggable binding site
A pocket on a protein surface that cannot bind with high affinity and specificity to small, drug-like molecules.

Decoy sites
A pocket on a protein surface that was identified using a computational algorithm but that does not correspond to the known ligand-binding site(s).

Druggability index
A prediction of a protein's capacity to bind with high affinity and specificity to small, drug-like molecules.

NMR hit rate
The frequency with which low molecular weight compounds from a fragment library bind to a protein as determined using 2D heteronuclear correlation spectroscopy.

and 'druggable' [6,7]. With the availability of the complete genome sequences for more than 100 prokaryotic and eukaryotic organisms (see www.ncbi.nih.gov/Genomes and www.tigr.org), we now have unprecedented access to large numbers of potential therapeutic targets. The question that arises is which specific protein targets will be druggable and disease modifying. It is estimated that ~10% of the entire human genome is involved in disease onset or progression [1], resulting in ~3000 potential targets suitable for therapeutic intervention. Added to these are the thousands of proteins from microbial and parasitic organisms, which could be targeted for the treatment of infectious diseases. Thus, the situation at the moment is not entirely unrelated to the dilemma illustrated in Figure 1. With a fixed budget and limited resources, only a fraction of these potential targets can be tackled at any given time. As it has been estimated that only several hundreds of the thousands of potential targets will be druggable and disease-modifying [1,7], the pharmaceutical industry is faced with a major challenge to rapidly and reliably identify the most promising targets for drug discovery efforts.

To discriminate between disease-associated versus disease-modifying genes, a variety of experimental and computational tools are being developed [8,9]. Using mouse models, large-scale genome-wide forward genetics screens (i.e. disease phenotype to gene), using either chemical mutagenesis (www.mouse-genome.bcm.tmc.edu/ENU/ENUHome.asp) or retroviral insertional mutagenesis (http://rtcgd.ncifcrf.gov), are employed to identify disease-causing genes. Conversely, large-scale reverse genetic screens (i.e. gene to disease phenotype) have proven to be successful in reproducing human disease phenotypes and are now utilized to identify therapeutically relevant targets that alter mammalian physiology [10]. The primary advantage of using animal models is the direct coupling of target identification and validation in an *in vivo* setting. Dramatic progress has also been made using *in vitro* cell-based technologies. The discovery of small interfering RNA (siRNA) has led to the development of high-throughput reverse genetic screens, in which specific genes are downregulated and the phenotype is assessed [11]. Large-scale transfection screens have also been performed to identify genes that elicit a transforming phenotype [12]. All of these methods are powerful in that they directly establish causality to the target. In addition to the above-mentioned functional approaches, genome-wide gene expression profiling has become routine in the identification of target candidates. This approach leverages alterations of gene expression levels in normal versus disease tissues. Genes with a high frequency of differential expression are likely involved in the disease process and might be amenable to therapeutic development. In the case of cancer, genome-wide DNA profiling is particularly informative, given that cancer-related genes often reside in recurrent genetic abnormalities [13].

Predicting protein druggability is another area that is experiencing tremendous growth. One approach for evaluating protein druggability is to analyze the genome on the basis of sequence homology to

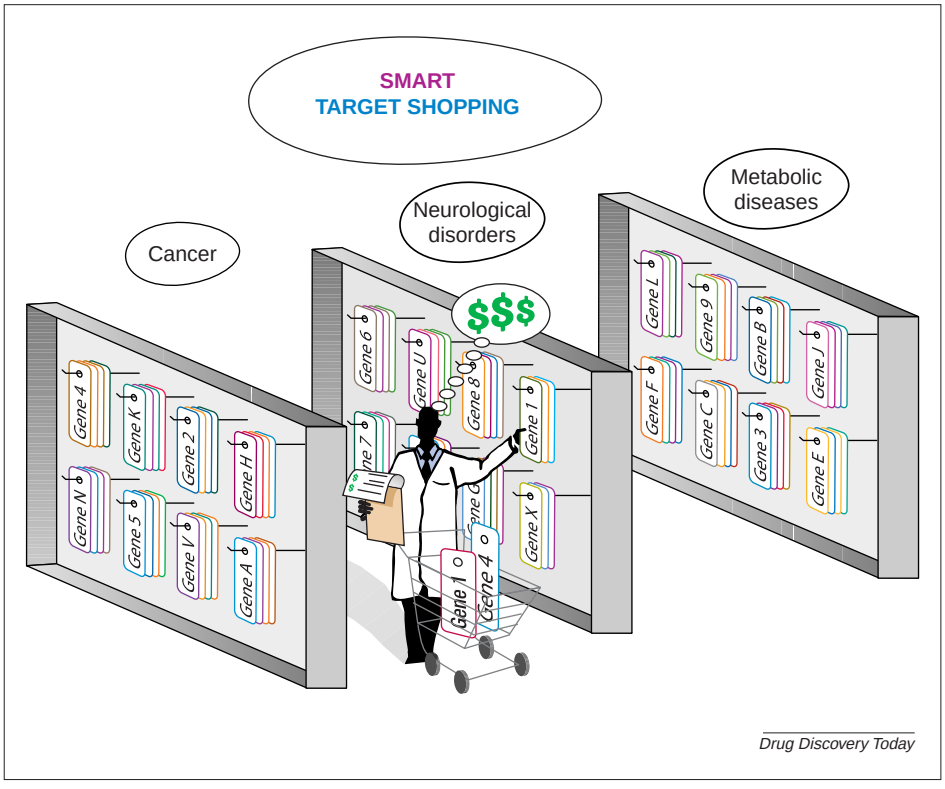

FIGURE 1

Target selection in the post-genomics era. With the complete sequence of the human genome in hand, there are now thousands of potential drug targets that can be pursued. However, only a fraction of these gene products are disease-modifying and vulnerable to intervention with small molecule therapeutics. With limited resources, pharmaceutical scientists must use all of the tools at their disposal to rapidly a reliably identify the most promising candidates.

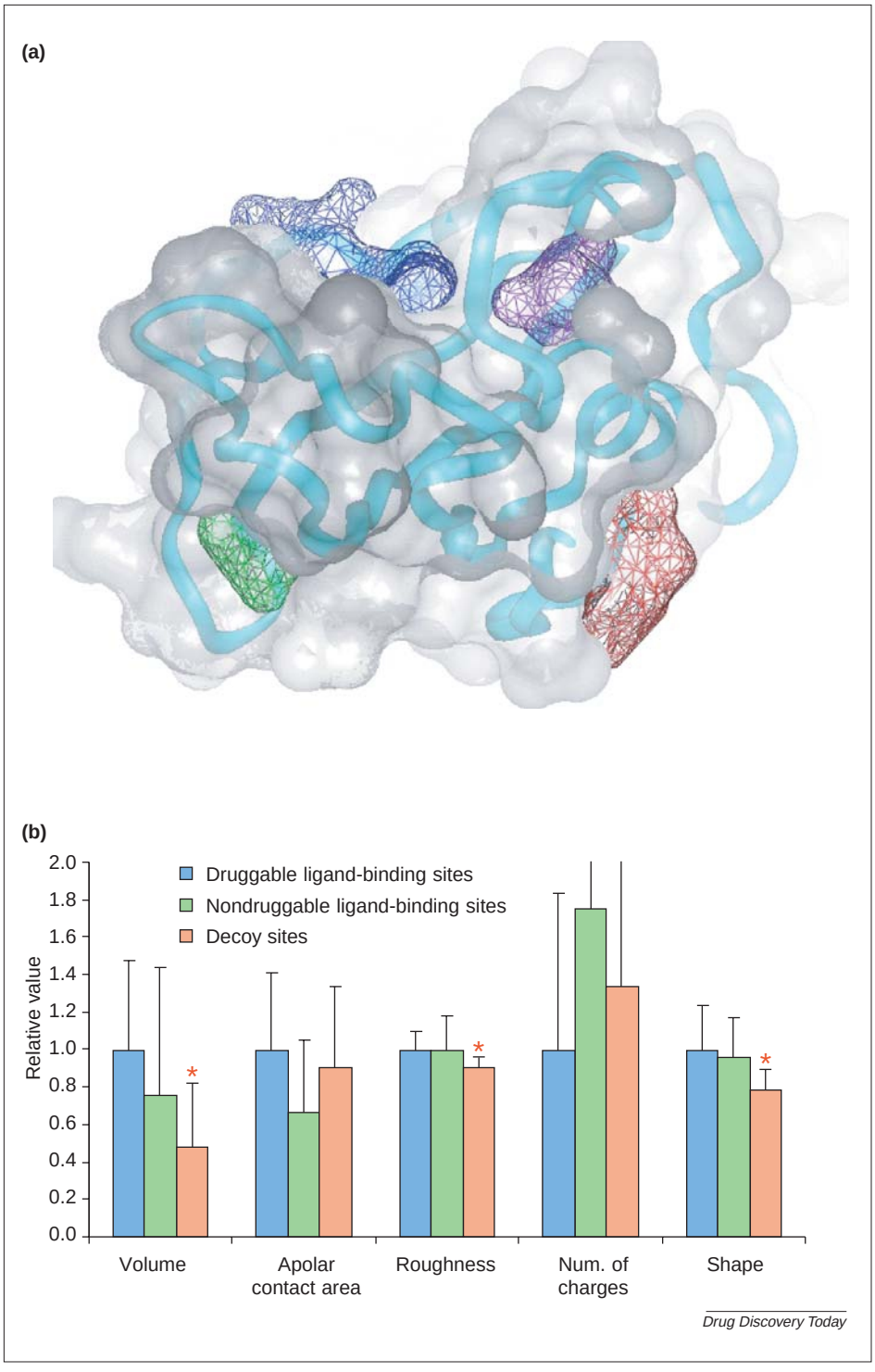

FIGURE 2

Identifying and characterizing protein binding sites. Geometry- or energy-based algorithms can be used to analyze a protein surface and identify potential ligand-binding sites. **(a)** Example of four protein pockets (shown in colored mesh) identified for the FK-506 binding protein (shown as a cyan ribbon with a gray surface) using the flood-fill algorithm available within InsightII (Accelrys). The sites depicted in blue and purple have been successfully utilized in the design of high-affinity ligands [37], whereas it has been experimentally shown that the red and green pockets do not bind to small, organic molecules [29]. **(b)** Comparison of several pocket parameters derived from 23 protein targets [29]. Druggable sites are defined as those for which the known active or ligand-binding site yielded NMR hit rates greater than 0.1%, whereas nondruggable sites yielded lower NMR hit rates. The decoy sites are additional pockets on the same set of proteins that are known not to bind to small, organic molecules. No statistical differences were observed in any of the parameters analyzed for the druggable and nondruggable ligand-binding sites, whereas the decoy sites were smaller, less geometrically complex and more elongated ($p < 0.05$, as indicated by the asterisks).

known therapeutic targets [1]. This bioinformatics analysis leverages an InterPro domain mapping strategy [14], which leads to an additional set of ~3000 unique therapeutic targets that comprise the 'druggable genome'. This includes G-protein coupled receptors, ion channels, protein kinases, metallopeptidases, proteases, nuclear hormone receptors, phosphodiesterases and other clinically validated target families. Of course, the pharmaceutically relevant targets are those that represent the intersection of the disease-modifying and druggable target sets. Thus, it is important to note that the current estimate for the druggable genome is based only on genes with annotated functions, which is approximately half of the entire human genome [1]. As such, additional classes of 'druggable' families might still be identified. Moreover, recent breakthroughs in small molecule disruption of protein–protein interactions (e.g. p53–MDM2 [15] and Bcl-2–Bax [16]) might lead to further expansion of the 'druggable' genome. As a result, assessing druggability simply by classify proteins into discrete target families could miss many protein targets that can be targeted with orally bioavailable, small-molecule drugs to achieve clinical efficacy in humans.

Predicting protein druggability

As described above, there is much research being conducted to identify protein targets that are disease-modifying and druggable. In addition to the many experimental tools to define molecular causation of disease onset and progression, new computational approaches are being developed that do not target specific protein families but instead rely only on the 3D structure of the protein target itself. Progress in this field is now enabling systematic analyses of protein surfaces in the search for binding pockets that have high potential to bind small, drug-like compounds with high affinity.

Finding protein pockets

The first step in trying to assess the potential druggability of a protein is to identify all possible binding sites on a protein surface (Figure 2). Many approaches for this have been described that can be broadly classified as either geometry-based or

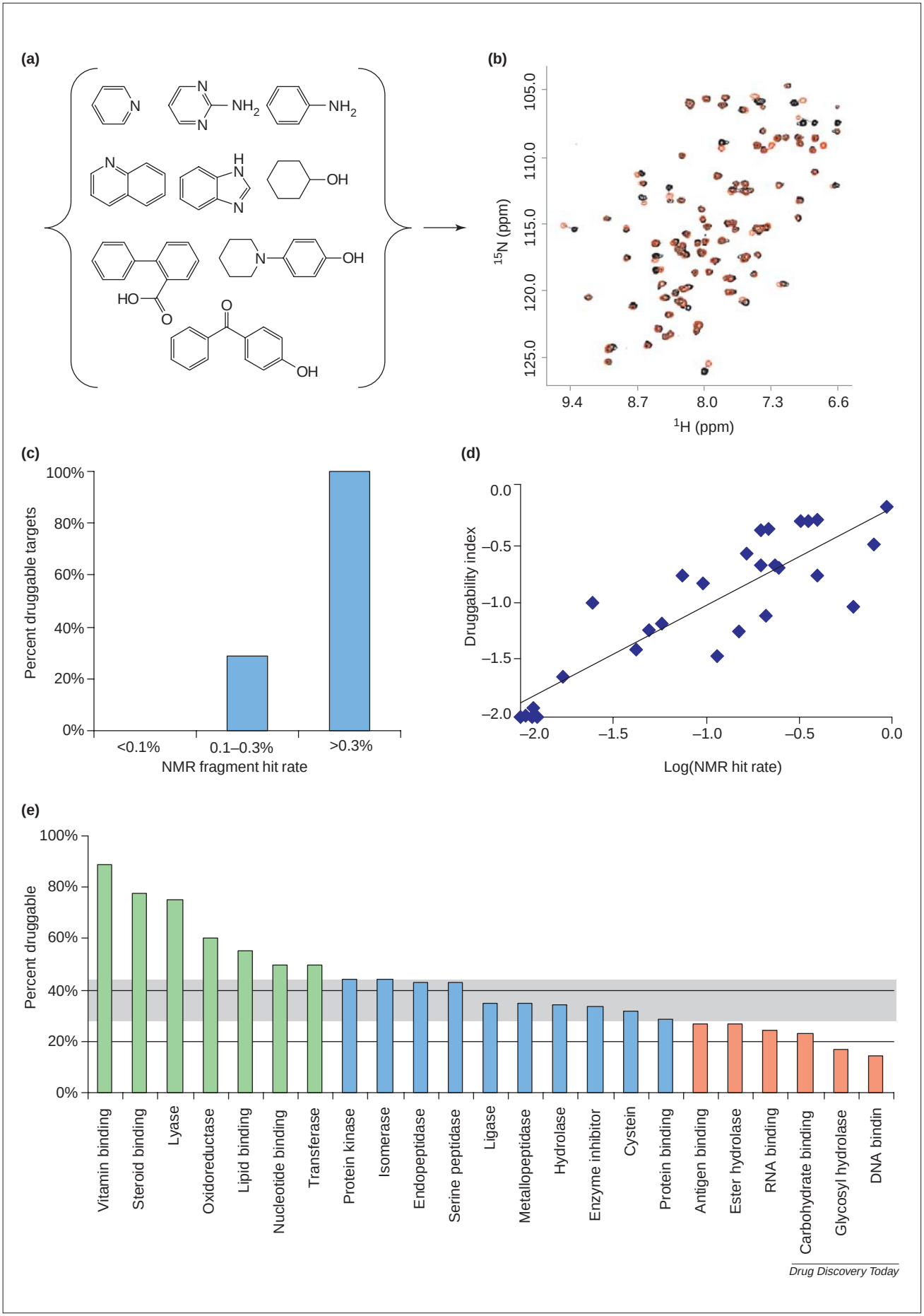
(a)
N
N
NH2
N
NH2
N
H
N
N
OH
N
OH
HO
O
O
OH
(b)
105.0
110.0
115.0
120.0
125.0
15N (ppm)
9.4
8.7
8.0
7.3
6.6
1H (ppm)
(c)
100%
80%
60%
40%
20%
0%
Percent druggable targets
<0.1%
0.1–0.3%
>0.3%
NMR fragment hit rate
(d)
0.0
–0.5
–1.0
–1.5
–2.0
Druggability index
–2.0
–1.5
–1.0
–0.5
0.0
Log(NMR hit rate)
(e)
100%
80%
60%
40%
20%
0%
Percent druggable
Vitamin binding
Steroid binding
Lyase
Oxidoreductase
Lipid binding
Nucleotide binding
Transferase
Protein kinase
Isomerase
Endopeptidase
Serine peptidase
Ligase
Metallopeptidase
Hydrolase
Enzyme inhibitor
Cystein
Protein binding
Antigen binding
Ester hydrolase
RNA binding
Carbohydrate binding
Glycosyl hydrolase
DNA bindin
Drug Discovery Today

FIGURE 3

Predicting protein druggability. Results from receptor-based NMR screening using a fragment library, in which a collection of low molecular weight compounds **(a)** is screened against an isotopically labeled protein [38–40]. The superimposition of 2D heteronuclear correlation NMR spectra **(b)** allows the detection of ligand binding. **(c)** The observed NMR hit rate is a reliable indicator of protein druggability. **(d)** Using these data, a correlation between the NMR hit rate and the apparent druggability of the protein could be derived, using simple pocket parameters. This algorithm allows for systematic and quantitative comparisons of protein pockets for potential druggability. Shown in **(e)** are the results of druggability indices derived from 1096 nonredundant human proteins, in which the percentage of members of selected target classes that contain a druggable binding site ('percent druggable') is plotted. In this preliminary analysis, no attempt was made to differentiate the known active or ligand-binding site, nor was conformational flexibility taken into account. Overall, 35% of the targets in this set contained at least one druggable binding site. Target classes depicted in green have a higher than average percentage of members with druggable binding sites, whereas those depicted in red have a lower than average percentage. The horizontal gray bar represents the average and standard deviation for the entire dataset.

energy-based algorithms. Geometry-based algorithms, such as SURFNET [17], LIGSITE [18], APROPOS [19], CAST [20], PASS [21] and the flood-fill algorithm available within Insight (Accelrys), all take advantage of the fact that natural ligand-binding sites tend to be concave invaginations in the protein surface. Another geometry-based algorithm calculates molecular surface complexity (called 'roughness') [22], with the premise that natural ligand-binding sites tend to be geometrically complex regions of the protein surface (Figure 2a). Intriguingly, in all of these reports describing the computational identification of protein pockets, the known ligand-binding sites were distinct from other surface cavities on the same proteins. For example, the ligand-binding site tends to be the largest [23] or most geometrically complex [22] cavity on the surface (Figure 2b). Energy-based algorithms, such as GRID [24], vdW-FFT [25], DrugSite [26] and computational solvent mapping [27,28], incorporate some level of physics into the pocket identification process by attempting to calculate binding potentials or binding energies. All of these approaches are successful in ranking the true ligand-binding site as one of the highest scoring pockets found on the protein surface.

Finding druggable pockets

Whereas finding the true ligand-binding site(s) is the first step in predicting protein druggability, the next step is assessing whether the binding site can bind with high affinity and specificity small molecule drugs. Unfortunately, few computational tools exist that can quantitatively assess a given binding site for its druggability. To capitalize fully on the information being generated by structural genomics initiatives, this ability is absolutely crucial. Although statistically significant differences in size and complexity exist between a true ligand-binding site and other sites on a particular protein surface, the picture is very different when the known ligand-binding sites between various proteins are compared. This is illustrated in Figure 2b, where various pocket parameters are charted for druggable binding sites, nondruggable binding sites and decoy sites (see Glossary). Interestingly, none of the parameters is suitable to differentiate between known ligand-binding sites that can or cannot be targeted with drug-like molecules [29]. Thus, although the known ligand-binding site might generally be the largest or most geometrically complex site on a given protein surface (as indicated by the statistically significant reduction in volume and roughness for the decoy sites shown in Figure 2), these parameters alone are not useful in discriminating between druggable and nondruggable binding sites between different proteins.

The most straightforward approach for directly assessing the druggability of a protein target is to actually perform a biochemical screen against a large compound collection and identify the number and types of small molecule hits. Presumably, those targets that yield multiple, structurally diverse hits are highly druggable, whereas those that do not yield any hit are nondruggable. However, one problem with this approach is ensuring that the compound collection is large enough to conclude that a target cannot bind small molecules. Given that the size of the chemical universe is estimated to be as large as 10^{60} compounds [30], even a compound repository of one million compounds reflects only a minute fraction of the chemical universe. In addition, deriving true hit rates from HTS data can be confounded by the often large number of false positives in the data [31]. To overcome these problems, we have recently reported on the use of NMR-based screening with a diverse fragment library to assess the potential druggability of protein targets in a reliable way [29]. The use of 2D heteronuclear correlation spectra (Figure 3b) is one of the most reliable methods for detecting ligand binding and yields false-negative and false-positive rates close to zero [32]. In addition, the use of a fragment library of low molecular weight compounds (e.g. see Figure 3a) reduces the possible size of the chemical universe by at least ten orders of magnitude [33,34]. Thus, an NMR-based fragment screen samples a significantly larger fraction of chemical diversity space and yields more reliable hit rates than conventional HTS. Using NMR-based screening, we observed a remarkable correlation between the hit rate from the fragment screen and the ability to produce high-affinity, drug-like leads for these proteins (Figure 3c) [29]. Significantly, some of the protein targets that were ultimately druggable yielded very low-hit rates from a conventional HTS of several hundred thousand compounds. In several of these cases, as suggested above, chemical matter simply did not exist in our corporate repository, with the necessary requirements for high-affinity binding to these proteins.

As an alternative to performing an NMR-based screen against every potential target, we were also able to derive 'druggability indices' from an analysis of the NMR data and the characteristics of the known ligand-binding sites (Figure 3d) [29]. Intuitively, it was found that characteristics such as charge, hydrophobicity, and shape were all

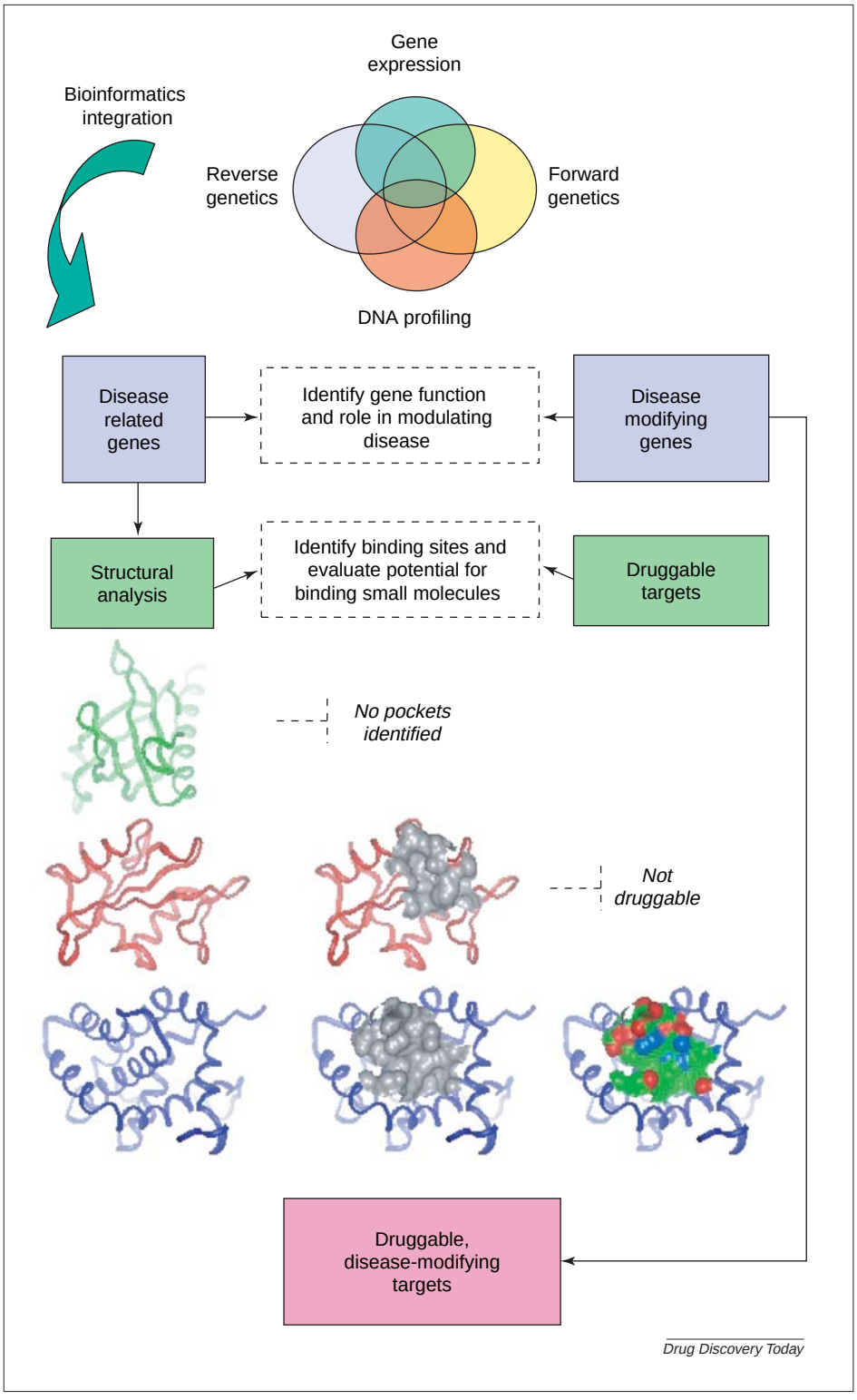

FIGURE 4

Incorporating druggability indices into target selection. Numerous target discovery databases can be analyzed to identify disease-related genes that are candidates for therapeutic intervention [e.g. reverse genetic approaches include transgenic knockouts and siRNA, forward genetic approaches include chemical and retroviral mutagenesis in mice, gene expression analyses include cDNA microarrays and serial analysis of gene expression (SAGE), and DNA profiling techniques include comparative genomic hybridization (CGH)]. These genes can then be analyzed using various molecular tools to elucidate gene ontology (the biological processes and molecular functions associated with a particular gene) and identify those proteins that are involved in disease onset and progression (disease-modifying genes). In parallel, structural analyses can be initiated of targets for which the structure is known to identify gene products that are likely to bind with high affinity to small organic molecules (druggable targets). In some cases, no pockets will be identified for a particular protein, suggesting that the protein itself is the ligand for another receptor. Proteins with identifiable pockets can be quantitatively assessed for potential druggability using the druggability indices described in this review [29]. Targets that are disease-modifying and contain pockets that are predicted to be highly druggable should be given highest priority for further discovery efforts.

important factors in predicting protein druggability. Importantly, however, the druggability indices derived from this work allowed an appropriate weighting of the various parameters so that quantitative comparisons of druggability could be made between different proteins. This allows for a systematic assessment of protein druggability given the 3D structures of the protein targets of interest. We have calculated the potential druggability of more than 1000 nonredundant human proteins derived from the Protein Data Bank (Figure 3e). In this preliminary analysis, nearly 35% of all included entries are predicted to contain at least one highly druggable binding site. Protein classes for which a large majority of the members presents druggable binding sites were vitamin and steroid receptors, with 89% and 77% of the members containing druggable pockets, respectively. Lyases (e.g. decarboxylases, carbonic anhydrases and aldolases) were also highly druggable, with 75% of the members containing druggable pockets. Other protein classes for which a higher than average percentage of members contain druggable pockets include oxidoreductases (60%), lipid- (55%) and nucleotide- (50%) binding proteins and transferases (50%). It is significant to note that, although 44% of protein kinases were predicted to contain a druggable pocket (higher than the average of 35%), the various conformations available to these proteins (i.e. from conformational changes upon phosphorylation and the high variability of several loop regions) immediately suggest the use of multiple crystal structures or the incorporation of conformational dynamics into the druggability assessment. Protein classes with the least number of druggable members were proteins that interact with DNA (14%), RNA (24%) and carbohydrates (17–22%). Surprisingly,

a substantial number of proteins involved in binding to other proteins (29%) contained druggable binding sites. This would suggest that a fair number of targets involved in protein–protein interactions might in fact be amenable to treatment with small molecules. Thus, the predicted druggability indices can be useful in prioritizing particular targets that would otherwise be avoided and to deprioritize targets that simply fall into a favored class.

Putting the tools together

As discussed in this review, there has been a virtual explosion of genomic and structural information that promises to revolutionize the way drug discovery is pursued. To maximize the potential for success, this information must be appropriately managed and analyzed so that the best target candidates can be rapidly identified, validated and incorporated into drug discovery programs. As illustrated in Figure 4, a bioinformatics approach could lead to the rapid identification of high value, therapeutically relevant targets. In this approach, meta-analyses of multiple genomic datasets (e.g. reverse and forward genetic screens, gene expression profiling and DNA profiling) would identify several disease-modifying genes that can be ranked on the basis of the frequency of the gene(s) across the datasets. Genes that are represented across multiple datasets are more likely to be disease modifying, given that the potential pitfalls of one discovery platform might be compensated by the strength of another platform(s). Candidate target genes might then be further categorized by gene function to prioritize those traditionally considered as druggable. Concurrently, the druggability indices of the genes that fall through the gene ontology filter could be directly assessed if sufficiently represented in the structural databases. Taken together, the utilization of this two-pronged approach would maximize the identification of therapeutically relevant targets that are disease-modifying and druggable, leading to an increased probability for successful drug discovery.

Conclusions

Finding drugs that are efficacious and safe continues to be a difficult, multi-faceted process. The ability to assess protein druggability in a fast and reliable manner is simply one of many tools that can help to streamline and enhance this process, especially when effectively integrated with other bioinformatic and experimental approaches to target identification and validation. That being said, it is important to realize that these computational and bioinformatic approaches will most certainly change and improve over the next few years. Bioinformatic scientists must identify optimal ways to merge the variable and disparate data that exist in the numerous genomic databases to capture the most promising candidate proteins. For predicting protein druggability, issues such as conformational flexibility will need to be addressed [35] and it can be envisioned that modified or even new pocket parameters (such as a more sophisticated treatment of electrostatics [36]) can enhance the predictive power of the algorithm. In addition, entire classes of protein targets are not adequately represented in the PDB (e.g. GPCRs), and breakthroughs in the structure determination of these important proteins will dramatically increase access to therapeutically relevant targets. Thus, whereas the current tools are reasonably successful in achieving their desired ends, it is anticipated that future developments will significantly improve our ability to rapidly and reliably identify druggable, disease-modifying proteins that have high potential for therapeutic intervention in humans.

References

1 Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730

2 Terwilliger, T.C. (2000) Structural genomics in North America. *Nat. Struct. Biol.* 7 (*Suppl.*), 935–939

3 Heinemann, U. (2000) Structural genomics in Europe: slow start, strong finish? *Nat. Struct. Biol.* 7 (*Suppl.*), 940–942

4 Yokoyama, S. *et al.* (2000) Structural genomics projects in Japan. *Nat. Struct. Biol.* 7 (*Suppl.*), 943–945

5 Michalsky, E. *et al.* (2005) SuperLigands - a database of ligand structures derived from the Protein Data Bank. *BMC Bioinformatics* 6, 122

6 Schneider, M. (2004) A rational approach to maximize success rate in target discovery. *Arch. Pharm. (Weinheim)* 337, 625–633

7 Betz, U.A. (2005) How many genomics targets can a portfolio afford? *Drug Discov. Today* 10, 1057–1063

8 Hardy, L.W. and Peet, N.P. (2004) The multiple orthogonal tools approach to define molecular causation in the validation of druggable targets. *Drug Discov. Today* 9, 117–126

9 Betz, U.A. *et al.* (2005) Genomics: success or failure to deliver drug targets? *Curr. Opin. Chem. Biol.* 9, 387–391

10 Zambrowicz, B.P. and Sands, A.T. (2003) Knockouts model the 100 best-selling drugs–will they model the next 100? *Nat. Rev. Drug Discov.* 2, 38–51

11 Xin, H. *et al.* (2004) High-throughput siRNA-based functional target validation. *J. Biomol. Screen.* 9, 286–293

12 Wan, D. *et al.* (2004) Large-scale cDNA transfection screening for genes related to cancer development and progression. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15724–15729

13 Albertson, D.G. *et al.* (2003) Chromosome aberrations in solid tumors. *Nat. Genet.* 34, 369–376

14 Mulder, N.J. *et al.* (2005) InterPro, progress and status in 2005. *Nucleic Acids Res.* 33 (*Database issue*), D201–D205

15 Vassilev, L.T. *et al.* (2004) *In vivo* activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303, 844–848

16 Oltersdorf, T. *et al.* (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 435, 677–681

17 Laskowski, R.A. (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph.* 13, 323–330, 307–308

18 Hendlich, M. *et al.* (1997) LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.* 15, 359–363, 389

19 Peters, K.P. *et al.* (1996) The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. *J. Mol. Biol.* 256, 201–213

20 Liang, J. *et al.* (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* 7, 1884–1897

21 Brady, G.P. and Stouten, P.F.W. (2000) Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Des.* 14, 383–401

22 Pettit, F.K. and Bowie, J.U. (1999) Protein surface roughness and small molecular binding sites. *J. Mol. Biol.* 285, 1377–1382

23 Laskowski, R.A. *et al.* (1996) Protein clefts in molecular recognition and function. *Protein Sci.* 5, 2438–2452

24 Goodford, P.J. (1985) A computational procedure

for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28, 849–857
25 Bliznyuk, A.A. and Gready, J.E. (1999) Simple method for locating possible ligand binding sites on protein surfaces. *J. Comp. Chem.* 20, 983–988
26 An, J. *et al.* (2004) Comprehensive identification of 'druggable' protein ligand binding sites. *Genome Inform. Ser. Workshop Genome Inform.* 15, 31–41
27 Dennis, S. *et al.* (2002) Computational mapping identifies the binding sites of organic solvents on proteins. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4290–4295
28 Kortvelyesi, T. *et al.* (2003) Improved mapping of protein binding sites. *J. Comput. Aided Mol. Des.* 17, 173–186
29 Hajduk, P.J. *et al.* (2005) Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* 48, 2518–2525
30 Bohacek, R.S. *et al.* (1996) The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* 16, 3–50
31 Diller, D.J. and Hobbs, D.W. (2004) Deriving knowledge through data mining high-throughput screening data. *J. Med. Chem.* 47, 6373–6383
32 Hajduk, P.J. and Burns, D.J. (2002) Integration of NMR and High-Throughput Screening. *Comb. Chem. High Throughput Screen.* 5, 613–622
33 Hajduk, P.J. *et al.* (1997) Discovering high-affinity ligands for proteins. *Science* 278, 497–499
34 Fink, T. *et al.* (2005) Virtual exploration of the small-molecule chemical universe below 160 Daltons. *Angew. Chem. Int. Ed. Engl.* 44, 1504–1508
35 Smith, G.R. *et al.* (2005) Incorporation of flexibility into rigid-body docking: applications in rounds 3-5 of CAPRI. *Proteins* 60, 263–268
36 Norel, R. *et al.* (2001) Electrostatic contributions to protein-protein interactions: fast energetic filters for docking and their physical basis. *Protein Sci.* 10, 2147–2161
37 Shuker, S.B. *et al.* (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274, 1531–1534
38 Hajduk, P.J. *et al.* (1999) NMR-based screening in drug discovery. *Q. Rev. Biophys.* 32, 211–240
39 Stockman, B.J. and Dalvit, C. (2002) NMR screening techniques in drug discovery and drug design. *Prog. NMR Spect.* 41, 187–231
40 Meyer, B. and Peters, T. (2003) NMR spectroscopy techniques for screening and identifying ligand binding to receptors. *Angew. Chem. Int. Ed. Engl.* 42, 864–890