Robert A. Stanley and William S. Hancock

Bioinformatics in the Clinic

CHALLENGES AND OPPORTUNITIES FOR IMPROVED TRIALS AND CLINICAL CARE

Abstract

As the potential values of bioinformatics applications in clinical settings are increasingly recognized, it becomes important to understand the state of the art. This paper describes key informatics challenges encountered during the final, information-intensive phases of the drug discovery and development (DDD) process—from preclinical and clinical trials to drug delivery, event reporting, and individualized patient care. Effective bioinformatics analysis tools should be capable of integrating comprehensive genomic, proteomic, metabolomic, and related data, including medical literature and patient demographic information. Current challenges to the application of informatics technology within the clinic include: overcoming barriers to connecting currently incompatible data distributed across different laboratories and points-of-care; keeping distributed data and analysis valid and audited; enabling efficient and secure report generation, report distribution, and data sharing; and providing the real-time analysis required for individualized treatment in distributed clinical environments. Current and emerging bioinformatics software methods for extracting meaning from data and for reporting within diverse molecular biological, clinical, and regulatory contexts are reviewed. Emphasis is given to technological challenges and opportunities for faster, more effective drug discovery and individualized treatment.

Key Words

Bioinformatics, informatics, proteomics, pharmacogenomics, systeomics, IND, NDA, clinical trials, individualized patient care, point-of-care, regulatory compliance, data integration, drug discovery and development, Sentient™, software

Fifty years ago, the discovery and description of DNA's double-helix nature by James Watson and Francis Crick signaled the beginning of a vital era for research and discovery, not only in molecular biology but also across the life sciences. The information-intensive DNA double helix is clearly a core structure within a complex network of interacting fields, each relevant to health, disease, evolution, even to life itself.

Today, genomics research interacts with complex new fields in molecular biology such as proteomics, metabonomics, and systeomics, each with complex data, multiple overlapping relationships, and analytical requirements that may at times seem too complex for the sharpest of human minds. Our ability to describe DNA has outpaced our ability to understand how this complex yet elegant molecu-

Mr. Robert A. Stanley *is Chief Intelligence Officer,* ***iSentient,*** *1325 61st St., Emeryville, CA 94608-2117, U.S.A.; tel.: 510-654-8256; fax: 510-654-8259; e-mail: rstanley@biosentients.com. Previously, Mr. Stanley served as Chief Information Officer at* ***Biosentients*** *(Emeryville, CA), an innovative bioinformatics software company and a division of* ***iSentient,*** *where he led the design of Sentient™ software products. Prior to founding* ***Biosentients,*** *he was employed as Associate Scientist by* ***Zeitgeist*** *(San Diego, CA), a software development company focused on bioinformatics, scientific software, and distributed computing. Mr. Stanley's employment at New York University's Bruner/Feldman Cognitive Science Laboratory and his advanced studies in epistemology, ontology, and applied network analysis have resulted in numerous conference presentations, publications, and invited memberships in the fields of epistemology, network software, distributed computing, and multidimensional analysis techniques.* ***Dr. William S. Hancock*** *is Bradstreet Chair in Bioanalytical Chemistry, Barnett Institute and Dept. of Chemistry, Northeastern University, Boston, MA, U.S.A. He received his B.Sc. and Ph.D. in Organic Chemistry, and later his D.Sc. in Biochemistry from the University of Adelaide (Adelaide, South Australia). Some of Dr. Hancock's previous positions include Vice President of Proteomics Development at* ***Thermo Finnigan, Inc.*** *(San Jose, CA), Principal Laboratory Scientist at* ***HP Laboratories*** *(Palo Alto, CA), and Director of Analytical Chemistry and Staff Scientist at* ***Genentech Inc.*** *(S. San Francisco, CA). Dr. Hancock is the recipient of several honors and awards, including the Walter Burfitt Prize and Research Medal from the Royal Society of New South Wales (Australia) for Pure or Applied Scientific Research; the Martin Gold Medal in Separation Science from the British Chromatographic Society (Nottingham, U.K.); the Csaba Horváth Research Medal from the Connecticut Separation Science Council (Storrs, CT); and, most recently, the Dal Nogare Award at PITTCON® 2003 and the American Chemical Society Award in Separation Science. He has served on several committees and is the author of 120 manuscripts accepted in international journals, 34 review chapters, seven books, nine patents, as well as over 160 invited presentations at international scientific meetings.*

lar structure is related to the "physiome"—the total organism.[1] The effective synthesis of the fields required to understand the relationship of the genome to the whole organism, and perhaps most importantly to enable application of our growing knowledge at the point of clinical care, demands new technologies.

Since 1953, the field known today as informatics has grown into a hugely important field in its own right. Specialized bioinformatics technologies are designed to integrate the molecular-level information contained within DNA, in the undoubtedly more complex universe of biological and clinical data, to enable unified analysis relevant to the complete, living organism. For example, current challenges are pressing for improved integration, management, and use of molecular-level data during the final information-intensive stages of the drug discovery and development (DDD) process, from preclinical and clinical trials to delivery, event reporting, diagnosis, and individualized patient care. Bioinformatics software platforms promise significantly improved acquisition, management, and analysis of global data relevant to molecular biology during these stages. However, it is widely known that more effective information management and bioinformatics methods than those currently in place must become widely implemented before the full value of distributed clinical data can be realized.[2] Conventional and emerging strategies in bioinformatics software will be reviewed in this article. Special attention will be given to historical challenges, conventional methods, and emerging strategies at the intersection of electronic information, molecular biology, and the clinic.

Opportunities for improved preclinical trials

The preclinical trials phase of the drug discovery and development life cycle often becomes even more laboratory focused than earlier phases, which often draw on data mining, computational modeling, and prediction assessed by ongoing experiments and assays. Even so, preclinical trials remain information intensive. After the researchers decide to move forward with a promising lead, the compound is tested in cell cultures, tissue samples, and laboratory animals (in vitro and in vivo). In practice, as few as 1 in 1000 initially promising lead compounds will actually make it through preclinical trials to the clinical phases.[3,4] Because most companies file patents for a promising lead during preclinical trials, patent protection for the drug begins to tick away while trials are still in progress. Preclinical trials average 18–36 months and can take up to 5 years to complete; any method to reduce this time and improve validation will provide clear value to the biotechnology and pharmaceutical industries as well as to academic researchers and clinical practitioners.

Challenges and opportunities in the manufacturing area also begin at this stage. Manufacturing offers a process and data management opportunity for bioinformatics software platforms, especially in the final drug delivery, clinical, and point-of-care stages. Manufacturing may be efficient and compliant with stringent regulatory requirements in order to succeed. The Waisman Clinical Bio-Manufacturing Facility (WCBF) at the University of Wisconsin, Madison, is one of many advanced biomanufacturing facilities that rely on paper for much of their data management and analysis needs. In many manufacturing scenarios, manufacturing parameters are generated from diverse instruments and require time-consuming and error-prone manual data entry and poorly validated cut-and-paste operations for regulatory and QA/QC reporting. Other software, or paper-related problems, can result from inventory management errors, loss of instrument parameter data, lack of real-time analysis of related manufacturing parameters, and failure to validate product purity profiles through persistent "always-on" monitoring.[5]

Currently, no company offers a commercial off-the-shelf (COTS) product that enables process validation and data acquisition and management for heterogeneous data and instrumentation across multiple networks. Products that provide this type of functionality will significantly improve manufacturing, inventory, and process management.

Challenges

Problems emerge from several avenues, such as barriers in functionally connecting preclinical laboratories (often located in separate locations or occasionally outsourced) with groups focused on analyzing results. Preclinical laboratory information must be effectively logged, normalized, validated, and integrated into the analytical environment. As additional in vivo and in vitro data are collected, new data types emerge that may not be compatible with the exploratory data and applications resources used previously.[6]

Laboratory information management system (LIMS) software from companies such as **LabVantage** (Bridgewater, NJ), **Thermo LabSystems** (Altrincham, Cheshire, U.K.), **LabLogic** (Broomhill, Sheffield, U.K.), and **Telecation** (Aspen, CO) can provide regulatory-compliant data histories and audit trails, but not for heterogeneous data from diverse database types, instruments, and applications. Few bioinformatics products provide the persistent data state management required to meet the FDA's Center for Biologics Evaluation and Research (CBER), Center for Biologics Evaluation and Research (CDER), and Code of Federal Regulations (CFR) 21.11 regulatory requirements for information analyzed across networks for multiple simultaneous users. None of the above tools provides regulatory-compliant any-to-any connectivity of instruments, data, applications, users, and user groups for the most flexible data integration and management configurations. An interactive environment that connects distributed data, instruments, and working groups *in a compliant manner,* even if they have different areas of expertise and prefer different computer operating systems and applications, is crucial at this phase.[4,6]

Emerging methods, products, and solutions

To make the critical decision to move from preclinical to clinical trials demands that all distributed information regarding success or failure of a lead compound be integrated, up to date, and managed in a validated manner. New bioinformatics and dis-

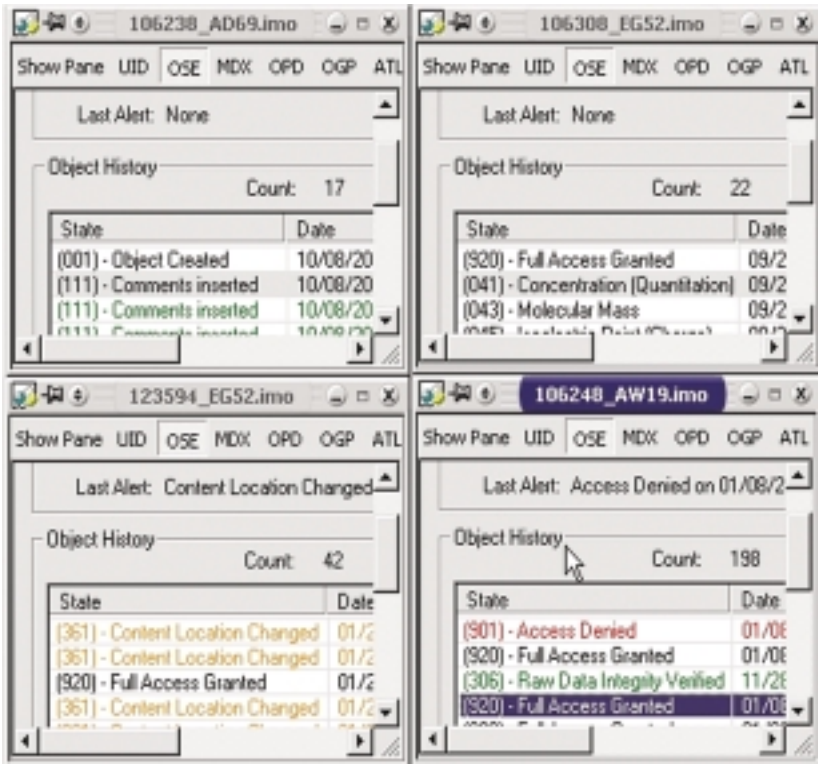

Figure 1 *Data state histories (e.g., record of status over time) for several Intelligent Molecular Objects™ (IMOs,* ***Biosentients,*** *Emeryville, CA). This type of functionality, coupled with selective security, will be crucial for software products that manage human clinical data.*

tributed LIMS-oriented methods are emerging to address these problems. "Intelligent" data structures, such as those shown in *Figure 1*, monitor data integrity, keep a detailed data state history for all analyzed data, and support compliant reporting in a distributed electronic environment. Using methods such as this, many of the current time-consuming cut-and-paste/paper trail steps for integrated, CBER/CDER-compliant preclinical and clinical validation and annotation can be automated, reducing the time and labor required to generate decisive, validated results.[7]

The bioinformatics software shown in Figure 1 keeps data history over normally "stateless" networks and maintains active data linking and tracking across networks. The Sentient™ platform technology (**Biosentients,** Emeryville, CA) uses an active object state engine to keep all state and processing history directly on the data and enable active linking and synchronization of scientific data and job flows across networks. Direct instrument acquisition and control methods allow access to data as they flow in from instruments. As new data are acquired, data property and activity history can be recorded for all data to enable persistent monitoring, validation, and ranking according to requirements for regulatory compliance.

Keeping data history and supporting references intact in this manner creates a powerful distributed database referring to all relevant data and required supporting information.[8] By providing immediate, unified analysis of results from distributed trials within a validated framework, tools such as these in the preclinical trials phase can cut the time required to complete effective preclinical analysis significantly. These methods also enable audited, selectively secure collaboration and data exchange across networks. In addition to improving efficiency in this stage, use of methods such as these will also improve the likelihood of success in the next clinical trials and new drug application (NDA) phases of the development life cycle. As they are integrated into the DDD process, bioinformatics software will enable earlier, more accurate, up-to-date answers to the expensive go/no go questions that become increasingly important as clinical trials continue.[7] Further down the line, these and similar methods will also increase the value of bioinformatics and clinical data for drug discovery and direct patient care.

Opportunities for improved clinical trials and new drug approval

If the preclinical data are promising, the company must decide whether to begin the clinical trials process. Since many biotechnology and larger pharmaceutical companies keep multiple drug candidates in the pipeline, the company must evaluate the size of the patient target market, competition in the market for that type of drug, and potential financial or other barriers to regulatory approval. If preclinical results and other information relevant to the decision-making process are significant enough to support the decision to move to human testing, the company applies to the FDA for Investigational New Drug (IND) status. Generally, clinical trials consist of three initial trials phases and a final NDA phase (*Figure 2*). Progress through the clinical trials phase requires at least 3–4 years in the current environment, and in some cases may take up to 9 years.[3,4]

Phase 1 clinical trials assess the drug's safety and basic activity profile, such as absorption, metabolism, effect, and duration of presence and activity within the human body. The sample size for Phase 1 ranges up to approx. 100 patients, usually comprising healthy volunteers. Phase 2 trials consist of tightly controlled experiments to evaluate safety and side effects, usually involving 100 or more volunteers suffering from the condition targeted by the drug. Phase 3 testing verifies the effectiveness of the drug and continues to build the safety and side effects profile over longer-term use. Phase 3 studies are normally double-blind studies with a sample of 1000 or more patients. If the drug proves to be effective in Phase 3, the trial is deemed successful. Occasionally, several repeated Phase 3 trials are required to ensure the validity of the studies; however, if the experimental conditions and validation

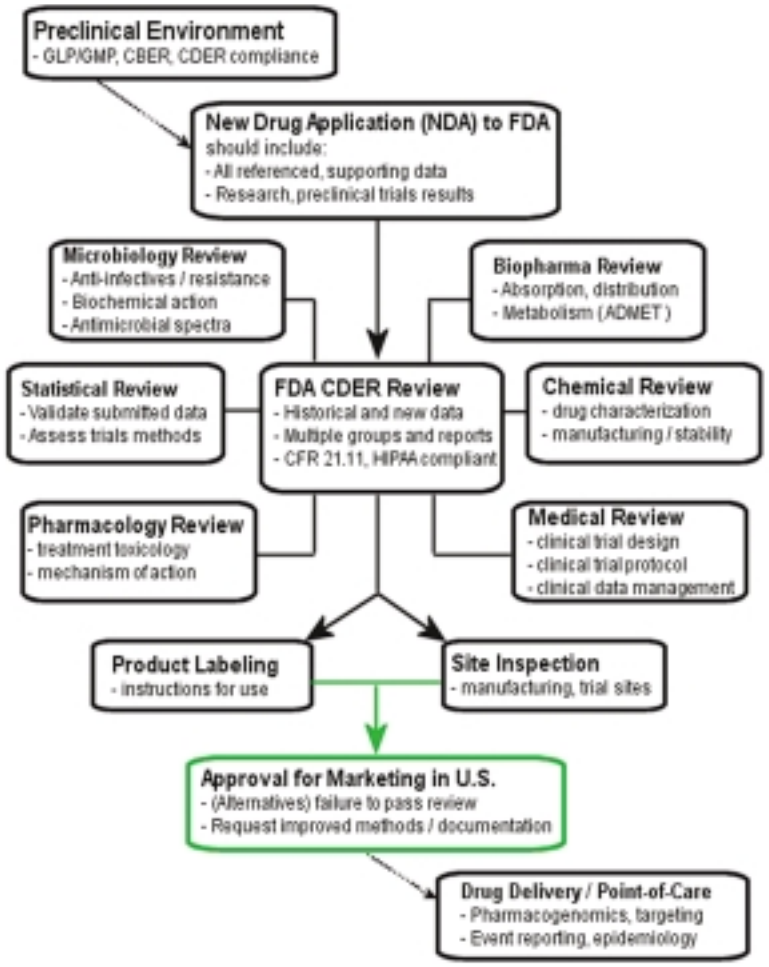

Figure 2 *Information-intensive NDA requirements. The NDA phase alone is extremely rigorous, demanding well-referenced reporting to several institutions in diverse formats. Paper documentation has been used traditionally with recognized and, should the FDA request further research or repeated trials due to inadequate documentation, very expensive shortcomings.*

are sufficiently rigorous and the results are sufficiently strong, one Phase 3 trial may suffice. Finally, if the IND successfully completes Phase 3 trials, the company files the NDA, which can contain up to 100,000 pages of validating data, in numerous report formats, for several responsible institutions.[3,4]

CDER provides strict reporting requirements in the Code of Federal Regulations for electronic records and signatures (21 CFR Part 11). If previous data acquisition and experimentation have followed CDER, CFR, ISO 900×, GLP, GMP, and GCP type validations requirements, and persistent data records and data integrity have been maintained, the reporting task will be facilitated.[5,6]

If software is in place to generate reports in compliant formats to the various institutions requiring experimental and clinical documentation, clinical trials may be completed very efficiently.

Bioinformatics tools used during the clinical trials phase must integrate existing and new clinical data and annotation in a validated format and efficiently generate the reports required to bring the new drug to market in a timely fashion.

Challenges

In order to achieve clinical status at all, the IND application must provide preclinical data and validated reporting of sufficient quality to justify human testing. According to Zeke Ashton's Cornell study,[3] even after meeting the stringent requirements to be granted IND status for clinical trials, promising leads have only about one chance in five of eventually reaching the market.

Clinical trials involve multiple, interactive data acquisition and reporting requirements for diverse clinical, research, and regulatory institutions. Thus, several problems face clinical trials phases. The lack of an interactive clinical interface capable of providing automated data integration from diverse clinical locations can restrict timely and uniform data acquisition. The absence of automated report generation in compliant formats slows down progress in meeting the clinical trials report requirements.

Companies such as **NuGenesis** (Westborough, MA) and **Valimation** (Paoli, PA) provide popular software products that support clinical compliance in a limited manner, such as for cataloging new data into a common database warehouse format (**NuGenesis**) or for regulatory reporting of clinical trials in multiple formats. Unfortunately, most popular LIMS and bioinformatics tools do not provide validated integration of heterogeneous, distributed data with all supporting references and processing history intact, coupled with the automated report generation.[9] For example, these products fail to provide methods to guarantee data integrity, enable persistent linking of key references to all experiments, and allow the selective security required to maintain the patient confidentiality required to carry out clinical trials over distributed networks.

The ideal bioinformatics tool will provide an interactive and unified interface for clinical data acquisition to permit validated, interoperable data output and reporting in real time. This interface should be intuitive and easy for the clinician to use, and provide data prepared for integration into validated report formats. Ideally, bioinformatics tools should automate the generation of reports required for all trial phases in fully compliant formats. Use of this type of technology will improve validation and significantly reduce the time required to complete clinical trials.

Emerging methods, products, and solutions

Efficient, audited interaction between research, clinical, and reviewing data groups and data and analytical resources will clearly improve clinical data feedback, refine drug and treatment targeting, and enable early

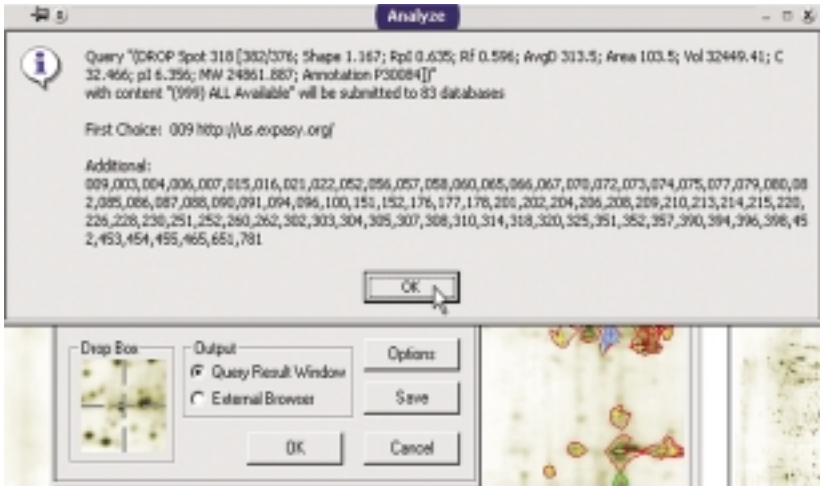

Figure 3 *Dialog box confirming parameters for a multidatabase query for all relevant information about a "spot" representing a protein biomarker from a human liver sample on a 2-D electrophoresis (2DE) data image. Depending on the data content and query profile, different databases are selected automatically from a list of ~150 heterogeneous databases, which may be public, subscription, or proprietary. Simple drag-and-drop capability, "plain English" prompts, and additional automated query profiling methods enable detection and alert for disease-related profiles indicated by comparison of the queried data to various genomic, proteomic, and metabolomic data resources. (Software by, and image reproduced with permission from,* ***Biosentients.****)*

intervention for improved results. **Biosentients** has developed an interactive interface for validated real-time reporting using the existing applications, instruments, and data resources required for research and trials. The Sentient platform provides a real-time interface to the universe of clinical data, prepared for integration into both research- and trials-related reports. Reports are customizable, with minimal programming required to refine reports for interactive communication with heterogeneous processes and institutions.[9]

To add to these challenges, it becomes even more important to integrate emerging public and private data and applications resources into the decision-making process at this stage.[8] Companies must continue to confirm that lead compounds remain patentable; keep track of competitive drugs in other company's pipelines or in the marketplace; and review all available scientific databases and publications to ensure that the compound remains a viable candidate or, conversely, is quickly removed from the DDD pipeline.

In addition to automated multidatabase queries, the Sentient platform provides data management and automated report generation in diverse regulatory-compliant formats (CBER, CDER, CFR 21.11, ISO 900×, CFIR, GLP, and cGMP). As the complexity of NDA phase reporting makes clear, tools to more efficiently meet requirements for validated electronic reporting will save human resources, time, and expense. Automated report generation based on secure, validated, up-to-the-minute data reduces the time required to successfully complete the clinical regulatory and NDA phase.

Together, these components ensure efficient clinical data acquisition; validated analysis; and secure, automated report generation based on comprehensive, audited database query methods (*Figure 3*). Technology such as this should improve the capacity of a company to move a promising lead through each of the regulatory trials phases and to market in a fast, well-validated manner.

Challenges and opportunities for improved delivery through clinical proteomics and pharmacogenomics

Validated real-time information management and integrated information systems will be recognized as crucial to the success of a drug even after the compound or agent is declared marketable by the FDA. Once a drug is brought to market, adverse events must be closely tracked and recorded and the target population for prescription and treatment should be optimized. Ongoing, real-time monitoring, data mining, diagnosis, and individualized care are all ideal components of this stage; they can be realized through the collaboration of bioinformatics with clinical proteomics, pharmacogenomics, and related disciplines. Optimized clinical delivery is crucial at this phase to maximize the benefit of the new drug in terms of clinical efficacy as well as financial viability. Real-time, individualized genomic and proteomic analysis (securely accessible over clinical networks) can provide the physician with the information required to detect otherwise asymptomatic disease and to prescribe the treatment most effective to a given patient.[10] This will improve disease detection, treatment prescription, and the benefit and viability of the drug product and treatment prescribed, and will minimize adverse events.

New information-intensive techniques are emerging to improve monitoring, profiling, modeling, targeting, and direct patient care throughout the full clinical systeomics field. These methods promise to improve, and in some cases enable, real-time epidemiology; clinical biocheminformatics; clinical proteomics; clinical absorption, distribution, metabolism, and toxicology (ADMET/TOX); population biosecurity; and pharmacogenomics.

Challenges

Bioinformatics-oriented software from companies such as **Spotfire** (Somerville, MA), **MDL** (San Leandro, CA), and **Oxford Molecular/ Accelrys** (San Diego, CA) can support customized process development, monitoring, and automation across networks. However, methods for integrating distributed data for real-time monitoring, diagnosis, and event reporting during this stage remain *ad hoc,* even when required by the FDA. Lost or incompatible data remain a problem in conventional clinical data management. However, tools for individualized pharmacogenomics and clinical proteomics offer an even more difficult scientific challenge and opportunity for contribution to the drug discovery and clinical domains.[11,12]

An interesting pilot for clinical proteomics is in progress at Northeastern University (Boston, MA), where the shotgun sequencing approach to proteomics is being used to look for marker proteins in individual patients with breast cancer. This approach uses laser capture microdissection (LCM) to isolate a specific cell type; then, the proteins in a small cell sample (typically 10,000 cells) are analyzed by HPLC coupled to ion trap MS. The measurement is performed on the corresponding peptides in an enzyme digest. Such a small sample, containing at least 10,000 proteins in a few micrograms of material, presents many challenges in avoiding contamination and losses of specific markers. An initial study[11] demonstrated the feasibility of this approach with a model cell line. Currently, the method has been extended to patient samples in a collaboration between Dr. Karger's group at the Barnett Institute and Dr. Palmer-Toy and Dr. Sgroi at Massachusetts General Hospital (Boston, MA). As noted in a recent issue of the *Journal of Proteome Research*,[10] "Clearly, the understanding of the molecular basis of any given disease would have tremendous impact on the development of new pharmaceuticals, ranging from speeding up the development process, to more rapid approval of new drugs, to the increased acceptance of a new therapy by a patient population. There is no better feedback than successful therapy in terms of new drug acceptance. So a worthwhile goal of proteomics in the future will be to develop molecular diagnostics for disease susceptibility and indicators of successful drug response. If personalized medicine comes to pass, then we will see a new behavior for the pharmaceutical industry, patients, and physicians—a valuable goal!"

Clearly, forward-looking research groups and consortia recognize the potential for valuable interaction among molecular biology, informatics, and the clinic, for example, in rapid growth in clinical proteomics and pharmacogenomics. As mentioned in Ref. 13: "Another anticipated tidal wave of change is in the use of pharmacogenomic information to streamline clinical trials. According to a recent survey conducted for the SNP Consortium, most companies believe that within five years, at least 50 percent of clinical trials will involve genotyping."

Leading bioinformatics software providers have yet to provide analytically integrated access to the distributed clinical data resources required for pharmacogenomics and individualized patient care. Integrated real-time access and automated analysis of currently incompatible data resources, ranging from broad patient demographics to genomic, proteomic, and pharmacological/chemical data, will be key to realizing the value of bioinformatics in the clinic.[13] It is crucial that research institutions, pharmaceutical companies, and clinical points-of-care alike optimize the value of newly marketed products by effectively manufacturing, monitoring, and profiling the entire drug delivery life cycle. Ideally, treatment parameters and clinical contexts associated with the drug will also be accessible and linked directly to all delivery activities.

Companies such as **Genaissance Pharmaceuticals** (New Haven, CT) are moving rapidly into the pharmacogenomics field. However, this and similar companies have yet to develop a bioinformatics system that provides real-time any-to-any connectivity and secure, validated access to terabytes of distributed data. This type of functionality is required for effective pharmacogenomics and individualized patient care, including interactive disease detection, drug/treatment targeting, adverse event reporting and reduction, and maximization of each drug's value to society, biotechnology, and the pharmaceutical industry.

Emerging methods, products, and solutions

To most effectively realize its potential benefit to science and society, bioinformatics technology must enable real-time integration and analysis of clinical data in the context of all relevant data resources.[7,9] Bioinformatics tools designed to allow comprehensive real-time distributed data access throughout the DDD lifecycle include next-generation bioinformatics platforms, which promise integration of preclinical laboratory data, comprehensive data resources, and the activities of collaborative preclinical research groups.

Figure 4 shows the Sentient platform depicting several of the main multidimensional relationships that can be analytically integrated to better understand disease characteristics, profile and validate drug activities, and detect contraindications and options for individualized patient treatments. Fast and efficient analytical integration of these normally incompatible data types, even when distributed across diverse databases and data types, is crucial for bioinformatics software and for biological research and discovery in the clinic.[13,14] In fact, this capability must be demonstrated before the promises of clinical proteomics, pharmacogenomics, and individualized patient care will be realized.

The lack of security appropriate to managing and exchanging human subjects' data over networks and across institutions provides another major challenge to clinical bioinformatics. As shown in Figure 4, precise data access and restriction methods can provide enhanced security. For example, depending on the user, entire patient records can be exchanged, patient data can be exchanged in an anonymous fashion, or small subsets of patient data can be sent for one or many patients at a time. This significantly improves compliance to security requirements, such as for data source validation, and patient confidentiality.

Security for analysis of distributed clinical information must be extremely well defined to meet patient privacy and validation; for example, identifying information should be selectively transparent (removed from some views, but not all).[7–9] This security is also unavailable to all but the most advanced bioinformatics products. While many open standards initiatives severely compromise security in the service of analytical integration, proprietary precise data access and routing techniques can provide integrated analysis with enhanced security.

Biosentients "intelligent data" methods provide an example of emerging proprietary software technologies designed to enable validated, real-time global data and applications integration off the shelf. Protected technology includes

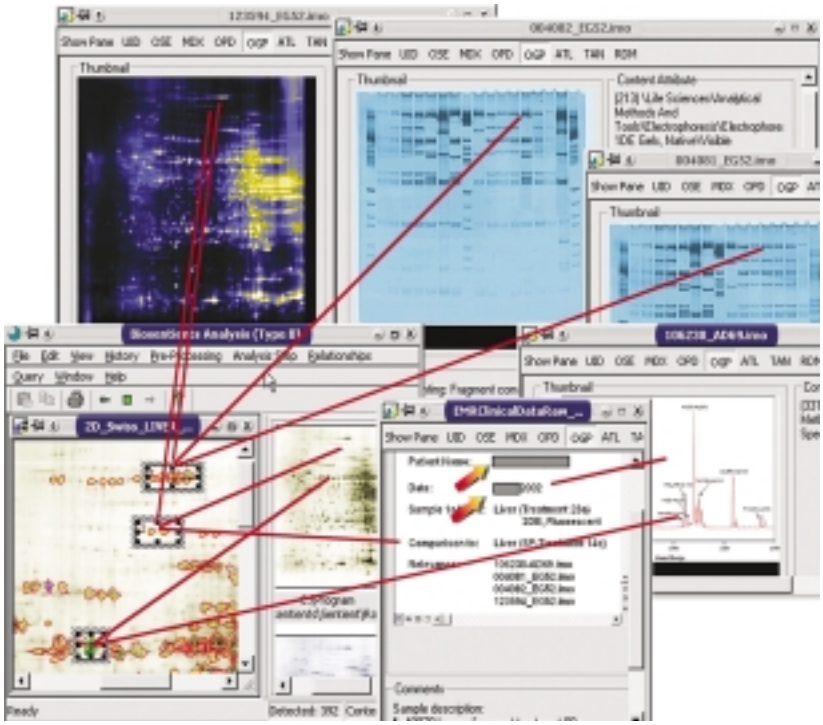

Figure 4 *Secure sharing and mining of preclinical and clinical data. Clockwise from top left: protein expression data (fluorescent 2DE image); genetic information about restriction fragment length polymorphism (RFLPs) on two related fluorescent 1-D electrophoresis (IDE) gel images, metabolic information (MS data), and confidential clinical records (arrows highlight information that may be selectively restricted or made accessible according to security level) in their queried relationships to protein expression information from a human liver sample (silver-stained 2DE gel image analyzed, bottom left). The bioinformatics software links vector subsets of data, such as those shown, to enable statistical analysis of property-based, functional, time-dependent, and other relationships. (Software by, and image reproduced with permission from,* ***Biosentients.****)*

direct-to-data vector subset input/output (I/O) methods that enable byte-level security; an active object state engine; and multiple interacting property panes that permit secure, integrated access to heterogeneous data across multiple operating systems, database back-ends, computing hardware infrastructures, and laboratory instrumentation.[7]

Opportunities for bioinformatics throughout the DDD life cycle

While bioinformatics software has historically failed to meet the regulatory and analytical demands required to effectively extract value from clinical data, emerging methods promise to radically increase the efficiency of clinical trials and the value of clinical data. However, innovative software methods are required to solve the problems described.

Placed within the undeniably meaningful clinical context, opportunities for improved drug discovery, delivery, population monitoring, diagnosis, and patient care are promising. Ambitious new technologies will enable integrated analysis and validated management of global clinical and laboratory data, literature, and all relevant supporting information. Products such as those reviewed promise to save years of expensive drug development time and expense, bring uniquely well-validated and targeted products to market, and usher in a new era of individualized patient care.[13,14]

In the context of comprehensive data and applications integration, any-to-any connectivity, security, and validated methods, emerging bioinformatics technologies will enable real-time population statistics, meaningful information management, and real-time event feedback at the point-of-care. This development promises significant improvement in functionality from trials to process management, metabolism and toxicology analysis, population monitoring, and individualized patient care. Bioinformatics solutions such as these promise to improve the rate and quality of drugs to market, minimize adverse drug events, and maximize patient benefit through pharmacogenomics and individualized bioinformatics.

References

1. Bassingthweighte JB. The physiome project: the macroethics of engineering toward health. The Bridge. National Academy of Engineering 2002; 32(3):24–9.
2. Wallace J. Personalized prescribing: advances in pharmacogenomics point toward a new era of individualized drug therapy. Scientist 2001; 15(12): 11.
3. Ashton Z. Clinical trials and the FDA. New York Hospital, Cornell University Medical Center 2000; www.nycornell.org/uo/fdatrial.html.
4. Hackett C. Regulatory affairs, path to approval. Reuters, Inc.: Armonk, NY, 1999.
5. Miller RA, Gardner RM. Recommendations for responsible monitoring and regulation of clinical software systems. J Am Med Info Assoc 1997; 4(6):442–57.
6. Fergus A, Kuehl D. Implications of 21 CFR Part 11 guidelines on the archival of analytical data. Am Lab 2003; 35(5):32–6.
7. Stanley RA, Gombocz EA, Stevenson R. Intelligent molecular objects: a new paradigm in bioinformatics. Am Genom Proteom Technol 2002; 2(4): 20–7.
8. Hancock WS, Wu SL, Stanley RA, Gombocz EA. The challenge of publishing large proteome datasets: the meeting of scientific policies and emerging technologies. Trends Biotech 2002; 20(12):39–44.
9. Gombocz EA, Stanley RA. Informatics: problems and solutions in the handling of massive amounts of disparate data. PharmaGenomics 2002; 2(2):30–40.
10. Hancock WS. Proteomic priorities. J Proteome Res 2002; 1(2):105.
11. Wu SL, Hancock WS, Goodrich GG, Kunitake ST, Hancock WS. An approach to the proteomic analysis of a breast cancer cell line (SKBR-3), in press.
12. Petricoin EF III, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002; 359:572–7.
13. Goble CA, Stevens R, Ng G. Transparent access to multiple bioinformatics information sources. IBM Systems 2001; 40(2):532–51.
14. Anderson NL, Anderson NG. The human plasma proteome: history, character and diagnostic prospects. Molec Cell Proteom 2002; 1:11:845–67.