

REVIEW

Informatics for neglected diseases collaborations

Frederic Bost*, Robert T Jacobs & Paul Kowalczyk

AddressSCYNEXIS Inc, PO Box 12878,
Research Triangle Park, NC 27709, USA
Email: frederic.bost@scynexis.com

*To whom correspondence should be addressed

Many different public and private organizations from across the globe are collaborating on neglected diseases drug-discovery and development projects with the aim of identifying a cure for tropical infectious diseases. These neglected diseases collaborations require a global, secure, multi-organization data-management solution, combined with a platform that facilitates communication and supports collaborative work. This review discusses the solutions offered by 'Software as a Service' (SaaS) web-based platforms, despite notable challenges, and the evolution of these platforms required to foster efficient virtual research efforts by geographically dispersed scientists.

Keywords Neglected diseases collaboration, public-private partnership (PPP), Software as a Service (SaaS)

Introduction

Tropical infectious diseases affect millions of individuals, predominantly in the developing world. The low financial viability for the sale of new pharmaceutical products in these poor countries does not offer an incentive to enable the high and risk-associated investments in R&D required for the discovery of new treatments for these diseases. Consequently, while these 'neglected diseases' affect many individuals, the research effort has been minimal for the past several decades [1,2]. Drugs currently used to treat these diseases are of limited availability and efficacy, are costly, and in many cases are based on old molecules, some of which have severe toxic effects. Furthermore, resistance to these drugs has emerged in several of these neglected diseases [3-5]. For the purposes of this review, the diseases considered as neglected are those diseases listed by the WHO [6] (eg, African sleeping sickness, visceral leishmaniasis, Chagas disease and schistosomiasis), in addition to malaria and tuberculosis. R&D efforts for these diseases are managed by virtual organizations, and thus information management represents a key challenge.

To address the need for better treatment of neglected diseases, several not-for-profit organizations have emerged in the past decade. These organizations are known as public-private partnerships (PPPs), as they share a common model that combines investment and expertise from the public sector and industry [7,8]. Examples of such PPPs include the WHO's Special Programme for Research and Training in Tropical Diseases (WHO/TDR) [4], the Drugs for Neglected Diseases *initiative* (DNDi) [9], the Medicines for Malaria Ventures (MMV) [10], the

Global Alliance for TB Drug Development (TB Alliance) [11], the Institute for OneWorld Health (iOWH) [12], the Malaria Vaccine Initiative (MVI) [13] and the Foundation for Innovative New Diagnostics (FIND) [14].

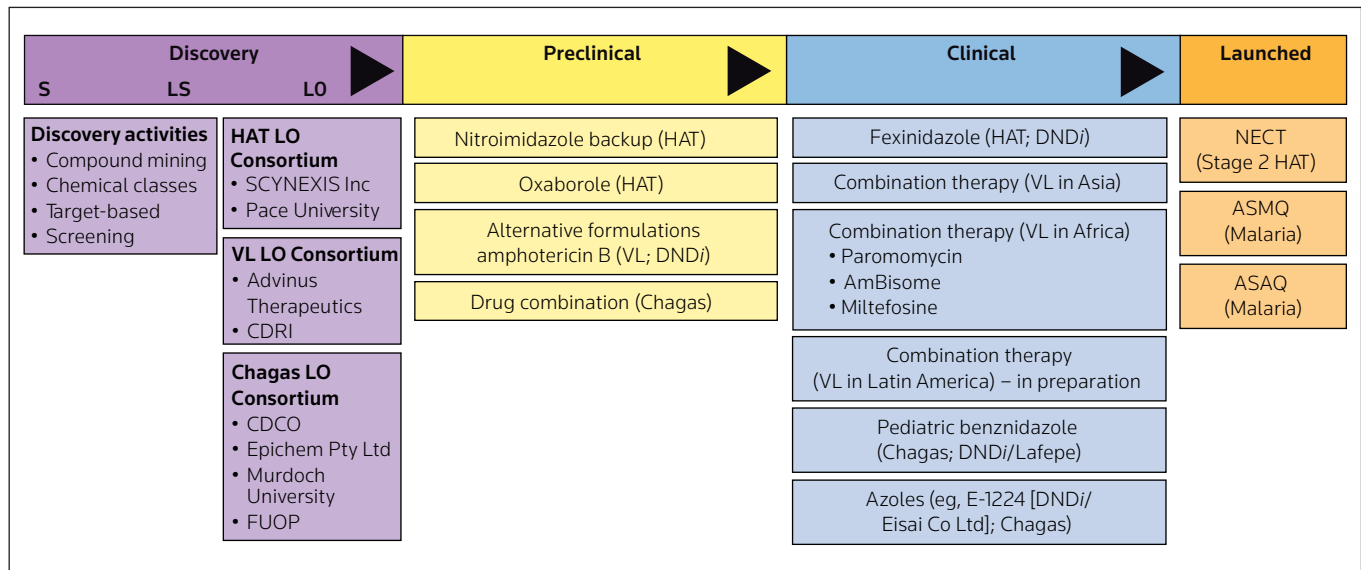
These PPPs have effectively raised money from different funding sources, including governments and charitable organizations such as The Bill and Melinda Gates Foundation and the Wellcome Trust. PPPs use these funds to conduct a virtual drug-discovery and development model, and have established a global network of multidisciplinary partners, including large pharmaceutical companies, biotechnology companies, not-for-profit organizations, CROs and academic institutions, in order to conduct R&D programs for various neglected diseases [4,9,10].

After having initially targeted the easy-to-achieve goals, such as expanding the applications of existing drugs [7,15], the PPPs have now evolved to address the entire cycle of drug discovery and development in order to sustain the delivery of innovative technologies. The PPPs coordinate multiple projects at every stage, managing a portfolio in a similar manner to large pharmaceutical and biotechnology companies [5,9]. An example of such a portfolio, which is managed by the DNDi, is presented in Figure 1.

The need for informatics tools to support virtual and global R&D collaborations

The visibility of the PPPs has impacted on the level of interest from biotechnology and large pharmaceutical companies in the area of neglected diseases, by

Figure 1. The DNDi's project portfolio.



A list of projects, as of January 2010, managed by the Drugs for Neglected Diseases *initiative* (DNDi) is presented. The DNDi manages projects at all stages of the development process, from discovery through to launched projects and is involved in exploratory research to create additional projects not represented in this figure. Major collaborators for projects in the discovery phase include GlaxoSmithKline plc, Anacor Pharmaceuticals Inc, Merck & Co Inc, Pfizer Inc, Novartis AG (eg, the Genomics Institute of the Novartis Research Foundation and the Novartis Institute for Tropical Diseases) and the Global Alliance for TB Drug Development. Sources for hit and lead compounds include the Eskitis Institute for Cell and Molecular Therapies and the Institut Pasteur Korea. In addition, the University of Dundee provides screening resources, and the London School of Hygiene & Tropical Medicine, the Swiss Tropical Institute and the University of Antwerp provide reference screening centers.

ASAQ Artesunate and amodiaquine combination therapy, **ASMQ** artesunate and mefloquine combination therapy, **CDCO** Centre for Drug Candidate Optimisation, Australia, **CDRI** Central Drug Research Institute, India, **FUOP** Federal University of Ouro Preto, Brazil, **HAT** human African trypanosomiasis (sleeping sickness), **LO** lead optimization, **LS** lead selection, **NECT** nifurtimox and eflornithine combination therapy, **S** screening, **VL** visceral leishmaniasis (kala-azar)

(Adapted with permission from the *Drugs for Neglected Diseases initiative*. © 2010 Drugs for Neglected Diseases initiative)

increasing awareness of both social responsibility within corporations and the strong need for the expertise of these companies in drug discovery and development [16,17]. By coupling the historical involvement of academic groups in research for neglected diseases with industry partners [18], it is now common for each PPP to manage collaborations with 30 to 40 different organizations worldwide.

The breadth and complexity of these neglected diseases collaborations require informatics solutions that facilitate and support such a virtual and global R&D organization. The first main challenge to be addressed is the traditional data integration from multiple scientific disciplines and contributors [19]. In the PPP model, another level of complexity is involved, as the data are spread globally, in multiple organizations with different cultures and processes [2,4]. Therefore, the informatics solutions must propose a resolution for the immediate and secure exchange of scientific information among the members of the network. Each result generated by a partner must be easy to share and consolidate with the other organizations regardless of location. In addition, an institutional memory that records the collaborative efforts should be generated in a central database, thus avoiding unnecessary duplication of effort [4,20].

Scientists involved in the collaborations require access to bioinformatics and cheminformatics software, both of which are traditionally used to perform data analysis and to help design future experiments *in silico* [21,22]. To permit remote access by any member of the partnership and to manage the diversity of environments, the software must be offered via a web-based front end and must be able to operate with all data captured by the network. In addition to data access and manipulation, the virtual nature of these projects requires informatics solutions that both enable and support teamwork between geographically dispersed members. Such solutions should include collaboration and communication tools that can manage the drug-discovery and development information, and project management and planning tools that allow the monitoring and prioritization of the projects portfolio, as well as the coordination of the partners [23]. Finally, and most critically, these informatics solutions must be delivered with strong security mechanisms in place to maintain the confidentiality of all information and to protect the IP that each partner has brought to the collaboration [3,4,24,25], as most of the neglected diseases collaborations do not evolve in an open-access or precompetitive environment [26,27].

Scientific requirements for drug discovery in neglected diseases

The scientific and data requirements for neglected diseases do not differ significantly from those used in traditional drug-discovery research, in particular the requirements directed at anti-infective or antimicrobial targets. Projects in drug discovery for neglected diseases can be classified either as 'target-based' – in which a biochemical target such as an enzyme, ion channel or receptor has been identified as of relevance to the viability of a parasite; or as 'phenotypic' – in which the ability of a drug candidate to kill the parasite via an unknown mechanism can be assessed in a whole-cell assay system. Both approaches require similar capabilities, but with subtle differences. Regardless of the approach taken, the main scientific requirements that allow projects to progress involve readily accessible, substructure searchable chemical databases and cheminformatics tools, which provide team members with the capability to explore widely diverse chemical structural classes for common features that correlate with a desired activity.

Target-based approaches

The emergence of genomic information for several parasites that cause neglected diseases, such as malaria [28], African sleeping sickness [29], leishmaniasis [30], Chagas disease [31] and cryptosporidiosis [32], has rapidly accelerated the ability to identify biochemical targets that are critical to the viability of these parasites. Disrupting the normal activity of such targets (eg, inhibiting a key metabolic enzyme such as trypanothione synthetase in *Trypanosoma brucei*) can be an effective strategy for the discovery of new drug candidates [33]. One important informatics capability required to support projects that use this approach is the availability of access to genomic information and the ability to develop correlations between antiparasitic activity and potential biochemical targets: this need has mainly been met by open-source tools [34-37].

In some cases, in particular for enzymatic targets, after a target has been identified and validated as essential to the viability of a parasite, modern molecular biology tools permit the cloning, expression, purification and crystallization of the protein target, potentially with an inhibitor bound to the active site. In these cases, the availability of contemporary molecular-modeling tools may be important to the discovery of compounds with improved efficacy [38,39].

Phenotypic approaches

A more traditional approach to the discovery of pharmaceutical agents for neglected diseases has been the evaluation of compounds directly on cultured parasites. As these assays do not require *a priori* information regarding the biochemical target, many of the bioinformatics tools used in target-based approaches are of limited value. However, after a class of compounds has been discovered to have a particular phenotypic

effect on a parasite of interest, a variety of image analysis tools [40] may be used to quantify these effects and rank the compounds in order of interest, as well as provide a method to compare chemical classes with one another. This approach may result in the discovery of common biochemical pathways, allowing for the transition from a phenotypic- to a target-based approach. Interestingly, for neglected diseases research, evidence suggests that this whole-organism screening approach can demonstrate a slightly reduced attrition rate for compound progression compared with traditional target-based approaches [4].

Collaborative web-based platforms for drug discovery and development

Few open-source software packages and/or databases are available that can satisfy the informatics needs of neglected diseases collaborations. Several public databases of genomic information, including the TDR Target and EupathDB databases, have proven useful for the bioinformatics community in the referencing of targets of interest for research in neglected diseases [34-37]. In the chemistry field, the PubChem and Low-Hanging-Fruit databases contain chemical and biological data relevant to neglected diseases research [41,42]. While valuable for providing access to publically available information, these databases are not designed to allow the secure and private hosting of data generated by third parties.

With regard to software, many free applications are available on the Internet [43]; however, accessing these applications from an open-source website is not advisable for confidentiality reasons. The level of adoption of such applications has been less obvious than the public databases [44], and the neglected diseases collaborations are not willing to sacrifice the quality of the software tools used, as illustrated by the vision of the DNDi, which states 'the best science for the most neglected' [45]. Commercially available software used by the pharmaceutical industry is costly and may not always offer a web-based front end to allow remote access by all users. When available, this software often lacks collaboration features and is not designed to allow restrictive access by multiple companies sharing a common database for the duration of a project.

As drug discovery and development has evolved to incorporate more outsourcing and collaborations, a new generation of software solutions has been developed, aiming to facilitate multi-organization partnerships. These solutions are web-based platforms proposed under a hosted 'Software as a Service' (SaaS) model. The SaaS platform is a fully web-based application that a user can access and conduct analyses via the Internet, but that is housed in the data centers of the vendor. These platforms require no software installation on personal computers, only an internet connection and browser. In the context of drug discovery, the SaaS platform is used to offer data-management

solutions via the Internet to geographically distributed partners. Each participant connects to the application from their location and, once authenticated, can upload scientific information to be shared with the project members, consolidate and organize results in the project database, search within all captured data and make informed decisions.

The two main examples of SaaS platforms for drug discovery are the Hit Explorer Operating System (HEOS) from SCYNEXIS Inc (see the *Example of a SaaS platform used by the neglected diseases collaborations: HEOS* section and reference [46]) and the Collaborative Drug Discovery Inc databases [47]. These web-based applications offer a cost-effective solution to connect multiple organizations via the Internet, as the vendor maintains the hardware infrastructure and provides consulting, training and support to all partners.

Challenges for the SaaS collaborative platform in drug discovery and development

The quality of service provided by SaaS platforms is critically dependent upon the ability of the provider to address certain key challenges successfully. First, the SaaS platform must propose solutions to exhaustively capture all information generated within the collaborations. The design of flexible solutions for data entry can be particularly challenging, as these solutions must accommodate all projects (eg, high-throughput screening campaigns or preclinical studies) and all types of partners, from large pharmaceutical companies to small institutes with limited infrastructure. The user friendliness of these registration tools is particularly essential for the acceptance of the platform and to avoid user frustration that could cause a reluctance to upload data by the collaborators.

Despite the large diversity of information involved, all information must be captured as efficiently as possible. For example, complex information related to the science of the project, such as chemical structures, physicochemical properties, NMR spectra, IC_{50} values and dose-response curves, *in vivo* reports and pharmacokinetic (PK) time-course studies, require different methods of data input (eg, textual, tabular or graphical) and file formats, but must be captured with high fidelity. Other information that must be captured is specific to individual partners and includes proprietary compound identifiers, which must be received through often diverse systems and infrastructures. The quality of the data captured is also of high importance, regardless of the origin, as decisions will be made based on the available data. As such, the SaaS platform must propose tools and workflows for standardization, harmonization, validation and quality assurance by a third party when required. A balance between the flexibility of the data entry tools and the necessity for standardization within projects that often perform the same type of biological assays must be generated [48].

After the data have been captured, the platform must offer easy access and robust functionality for all partners. While web-based interfaces facilitate data access, all types of scientific data searches must be possible, including chemical substructure searches. The software architecture must be designed to satisfy scientists monitoring the process at several different levels, and to allow for the data to be grouped in a flexible manner. For example, an individual scientist may need to view data at the level of the container barcodes, whereas a team leader building a SAR table may require grouped data, and a project manager may measure the outcome of the network and benchmark projects across broad descriptors such as chemical series or testing location.

Finally, a critical point is the security of the SaaS platform, which must rely on a highly secure infrastructure at the physical, network and application level. In addition, the security scheme must offer advanced granularity in order to set up the data and functional access rights. Providing the PPPs with the ability to segregate projects and allowing individual project members to see only project-specific data is essential, but not sufficient, to cover the variety of user profiles and processes that must be addressed. It is also essential to design and communicate the authorization processes that allow access to be granted. Only the reliable implementation of all these factors will provide the assurance required by collaborators, in particular large pharmaceutical companies, to deposit data for use in projects. Otherwise, project participation can be limited, and the removal of data can occur, thereby rendering other data less meaningful.

Example of a SaaS platform used by the neglected diseases collaborations: HEOS

HEOS was developed initially to support a contract research business 7 years ago, and has been an effective solution for neglected diseases collaborations. During the past 3 years, HEOS has hosted data for several major PPPs, including the WHO/TDR, the DNDi and the MMV [9,21,49]. As such, HEOS manages the information generated globally by more than 400 users in 60 organizations involved in neglected diseases discovery and development (Figure 2), including large pharmaceutical companies, CROs, biotechnology companies, non-profit organizations and universities. HEOS hosts more than 90 drug-discovery and development projects, 600,000 well-characterized chemical compounds, 1.5 million validated biological data points, 200,000 validated drug metabolism and PK results and 5000 project documents.

The DNDi-SCYNEXIS human African trypanosomiasis (HAT) program (Figure 1) is an example of a drug-discovery project in neglected diseases that has benefitted from HEOS. Initiated in 2006, this program involved a high-throughput screen of approximately 25,000 compounds for their ability to kill *T brucei*, the causative parasite of HAT, and has progressed through

Figure 2. Geographical representation of organizations connected to HEOS.

As of January 2010, 60 organizations involved in neglected diseases drug discovery and development were connected to the Hit Explorer Operating System (**HEOS**) developed by SCYNEXIS Inc.
(Adapted with permission from SCYNEXIS Inc. © 2010 SCYNEXIS Inc)

hit-to-lead and lead optimization phases; a preclinical candidate, SCYX-7158 (Anacor Pharmaceuticals Inc/ Drugs for Neglected Diseases *initiative*/Pace University/ SCYNEXIS Inc) has now been selected [50]. As this project evolved, geographically distributed partners from industry, government and academia have been engaged to provide experimental capabilities and expertise (Figure 3).

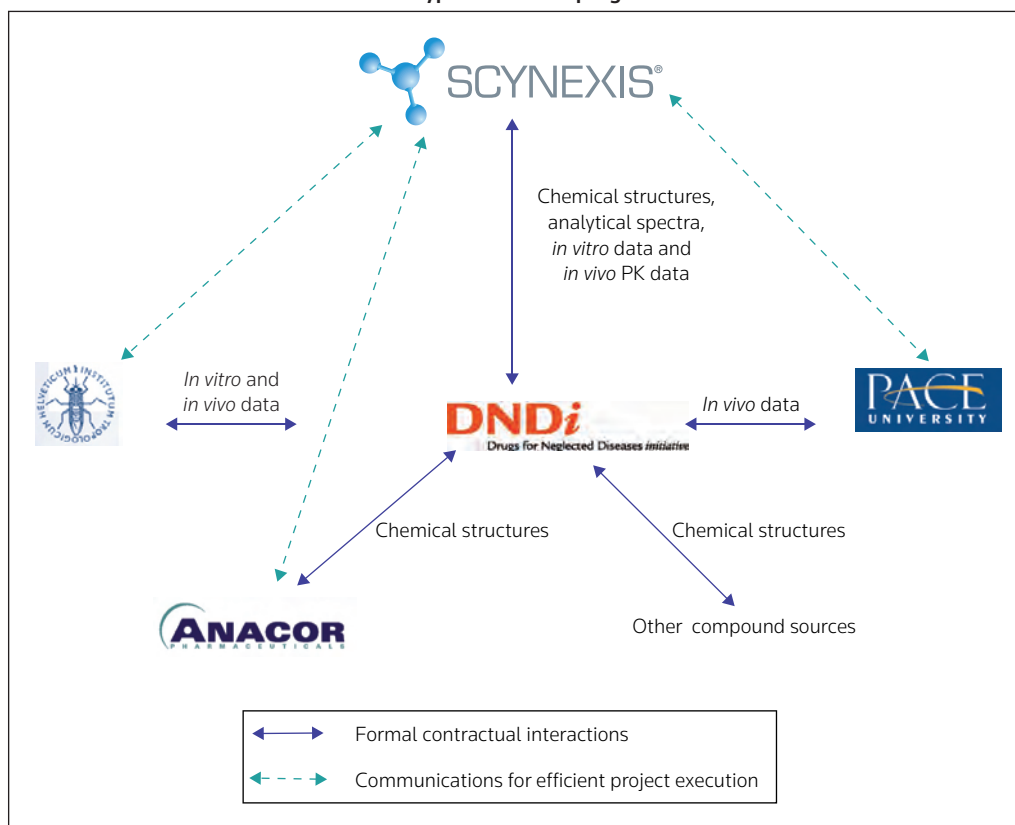
The various partners involved in the HAT program have provided and/or accessed different information in diverse formats (Figure 3). For example, SCYNEXIS has been responsible for the design, synthesis and primary screening of new drug candidates and PK screens; Pace University provided an *in vivo* evaluation of drug candidate efficacy; and the Swiss Tropical Institute confirmed activity against additional parasite strains. In addition, the DNDi has actively pursued the acquisition of novel chemical leads from a broad range of collaborators across industry and academia. Data generated on these compounds by one of the screening centers were also stored and made available to the core team. An example of an acquired chemical series is the organoboron chemotype initially discovered by Anacor Pharmaceuticals and optimized for HAT within the project by SCYNEXIS.

Because of the diversity of data input and information output within the project, considerable flexibility has been required of the SaaS platform. Data input from

high-throughput screening uses an automated upload of transformed data (via ActivityBase [51]) directly from the instrument on which the data are collected. Conversely, data from secondary *in vitro* and *in vivo* screens (eg, efficacy and PK data) are manually curated by the researcher and inputted via custom-designed Excel templates. With regard to output, tools within the platform allow the user to export numerical data in several tabular formats for use in statistical analysis packages, whereas more complex graphical data are transformed into standard image files to facilitate inclusion in reports and presentations. An underlying capability important to all users is the chemical knowledge contained within the database, which is inputted by researchers through chemically aware registration and substructure searchable query features. Figure 4 provides an example of a screen used to access all the captured data for a given compound.

As discussed in the *Challenges for the SaaS collaborative platform in drug discovery and development* section, the ability to provide secure access to authorized users associated with the various collaborators has been a key component of HEOS. As project sponsor and coordinator, the DNDi has full access to all data on all compounds, but individual collaborators, such as SCYNEXIS and Anacor Pharmaceuticals, have access only to compounds and data as authorized by various agreements and IP considerations. All access has been customized at the

Figure 3. Overview of the DNDi-SCYNEXIS human African trypanosomiasis program.



The collaborations between the various partners involved in the Drugs for Neglected Diseases *initiative* (DNDi)-SCYNEXIS Inc human African trypanosomiasis program are depicted.

PK Pharmacokinetics

(Adapted with permission from SCYNEXIS Inc. © 2010 SCYNEXIS Inc)

user level as the project has evolved, with access granted and removed at the discretion of the project sponsor.

Turning data into knowledge

While curated data repositories are necessary for any drug-discovery research, it is only when these data are mined, visualized and analyzed that a knowledge base for a particular disease target can be built. Data analyses range from the removal of compounds with reactive groups from a compound library, to measuring rule-of-five compliance to ligand-based and structure-based modeling. In addition, as robust screening data becomes available, the opportunities to build SARs become more viable.

The workflow technology embodied in Pipeline Pilot (a scientific informatics platform from Accelrys Inc) [52], lends itself to the facile execution of several cheminformatics tasks, including, but not limited to, substructure searching, clustering, the construction of Bayesian models, combinatorial library enumeration and the numerical characterization of compounds for further statistical analyses.

Tibco Software Inc's Spotfire Analytics provides a platform for visual analysis that allows an end-user to interact with data in real-time [53]. This data visualization is particularly useful for interpreting relationships in multidimensional data. Interactive filters further aid the data analysis workflow, and end-users can quickly identify compounds of particular interest within a data set. Furthermore, the end-user can also construct 'what-if' scenarios; for example, to consider the consequence of restricting the maximum molecular weight to 400 Da, the maximum number of rotatable bonds to five and the maximum number of hydrogen-bond acceptors to four.

The R Foundation has developed R, a language and environment for statistical computing and graphics [54] that includes a large, coherent and integrated collection of intermediate tools for data analysis, as well as graphical facilities for data analysis and display. Statistical techniques available through R include linear and non-linear modeling, classical statistical tests, classification and clustering.

Pipeline Pilot, Spotfire and R are all integrated within HEOS, affording end-users access to state-of-the-art data analysis, visualization and mining tools (Figure 5).

Figure 4. HEOS single-compound view.

The screenshot shows the HEOS single-compound view interface. The main window displays the following information:

Compound ID: SCYX0000902917
Compound Name:
Trivial Name:
Exact Mass: 329
Molecular Weight: 329.443993
Molecular Formula: C₁₉H₂₇N₃O₂
Stereo Designation:
Status:
Status Comment:

Calculated Physchem Properties

Assay Results Table:

Sample	Assay Type	Organism	Test Parameter	Dose	Test Value	Test Date	Tester
2343847	Malaria in vitro	P. falciparum	IC50		= 0.0022 ug/ml	01-MAY-2007	Balela Nare
2343847	hAT in vitro	T. b. modesiense	IC50		= .003 ug/ml	01-MAY-2007	Balela Nare
2343847	CYTOTOXICITY	MRC-5 cells	IC50		> .055 ug/ml	01-MAY-2007	Balela Nare
2343847	Malaria in vivo	P. berghei	% inhibition	30 mg/kg	= 45 %	01-JUN-2007	Balela Nare
2343847	Malaria in vivo	ANKA (p)	% inhibition	4 adx days	= 30 %	20-JUN-2007	Balela Nare
2343847	Solubility	PBS - pH 7.4	Kinetic		= 30 uM	15-MAY-2007	Bernard Soomeux
	ADME	ADME_IT					

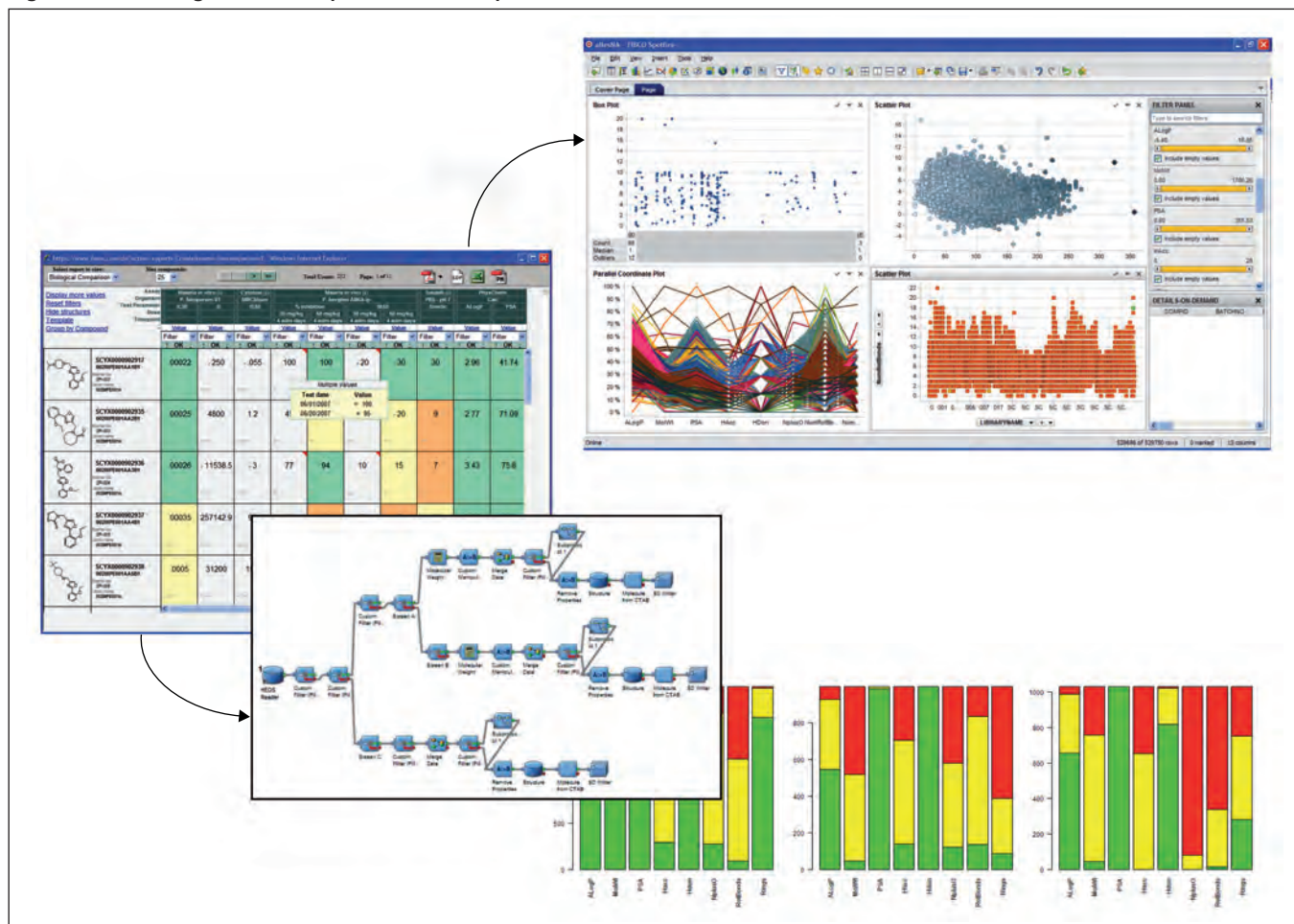
The single-compound view from the Hit Explorer Operating System (HEOS) allows access to all the information captured for a given compound.

Collaboration and communication tools

In addition to data and information sharing from a database perspective, tools that facilitate the collaboration and communication of researchers in the neglected diseases network are also required. Some collaborative tools have already demonstrated success in the drug-discovery field: examples include the use of Microsoft Sharepoint within the firewall of a large organization such as Pfizer Inc [55], or the use of open-data sources such as UsefulChem [56]. Such tools not only contribute to a better collaboration, but also capture a new wealth of information because these data are searchable (eg, the rationale behind the design of NCEs). Notably, the eRoom capabilities of HEOS are used extensively by multiple project teams, thus enabling the knowledge captured within the shared documents (ie, reports and presentations) to be at least as valuable as the data stored in the database. One example includes a clinical project that has been driven by partners distributed globally, using a formalized folder structure in the eRoom.

Another set of promising tools are those currently developed by the consumer market to facilitate communications between individuals. Facebook, LinkedIn, Twitter and Google Wave are some of the new web services that allow instant communication between contacts in several different manners and the exchange of any type of media [57-60]. These tools provide the potential for increased communication within the network, and some drug-discovery public communities have already emerged within Facebook [61] and LinkedIn [62]. However, one limitation of these tools is the inability to handle scientific data efficiently, but it is encouraging to note that some vendors have already started to address this limitation. For example, the collaboration between Accelrys and Microsoft to enable SharePoint [63], CambridgeSoft's addition of chemical capability to the collaborative platform of Imaginatik plc [64], Symyx Solutions Inc's ChemMobi [65] and ID Business Solutions Ltd's ChemJuice [66] for the iPhone. The expansion of these capabilities to other devices, such as the smartphone, will further facilitate communication and provide instant access to the SaaS drug-discovery platform.

Figure 5. HEOS integration with Pipeline Pilot and Spotfire.



Data captured in the Hit Explorer Operating System (HEOS) database can be sent to Tibco Software Inc's Spotfire Analytics (top right) and Accelrys Inc's Pipeline Pilot (bottom right) for advanced data analysis.

The extension of these capabilities to collaborations involving confidential information from multiple companies will require a platform that is both secure and private to host these services. Such an extension may be developed by the SaaS drug-discovery platform to offer a tightly integrated solution alongside the data-management capabilities. This approach has recently been investigated by the pioneer SaaS company salesforce.com, through the development of Chatter, the company's own embedded capability [67]. Currently used tools may also evolve, as the industry beyond drug discovery is likely to be willing to adopt confidential versions [68]. The SaaS drug-discovery platforms would then have to integrate these versions to provide a combined service.

One distinct benefit of these communication tools is the development of a sense of community [57], matching the intrinsic nature of the neglected diseases area in which competition to achieve market share is minimal because of the nature of the diseases. After the security aspects inherent in the drug-discovery industry have been addressed, scientists can share daily experiences,

opinions and commentaries within the project, thus a higher adoption rate of the technology and the creation of many 'semi-private' neglected diseases communities can be expected. Such a vision was expressed during the creation of the African Network for Drugs and Diagnostics Innovation [69].

Conclusion

Neglected diseases collaborations consist of many different organizations working together to find cures for the diseases of the world's most neglected individuals. The ability of these collaborations to combine public and private expertise and investment have contributed to several success stories (eg, see references [15,18,70]).

Because of the geographical distribution and virtual nature of these collaborations, the support of customized informatics solutions can have a significant impact. The potential of these solutions has led to the early adoption and evolution of the SaaS drug-discovery and development web-based platforms by this community. Because of these platforms, several major PPPs now control

data management for various collaborations. As these platforms grow and are further interfaced with powerful data-analysis tools, scientists will have a greater ability to convert data into knowledge. These SaaS platforms have evolved to provide solutions similar to the ones developed and used within the firewall of large pharmaceutical and biotechnology companies.

The evolution of the Internet and the wealth of new tools such as Twitter and Facebook could soon offer significant communication enhancements if adapted to handle secure and private scientific communities. The neglected diseases collaborations that have already pioneered new types of hosted solutions are now well prepared to evaluate any novel tools that the industry delivers. If a certain tool fills a gap, these collaborations are in a position to adopt this tool rapidly, deploying it among their partners to generate immediate benefits.

The combination of secure and scientific services offered by the SaaS platform for experimental data management, data analysis, remote collaborations and advanced communication can significantly aid virtual team members in the building and leveraging of knowledge shared among the organizations, and thus increase the chances of finding new treatments against diseases affecting the most neglected.

Acknowledgements

The authors would like to thank Terry Marquardt (SCYNEXIS), Dominique Besson (WHO/TDR), Jean-Robert Ioset and Federica Giovannini (DNDi), and Pascal Fantauzzi and Jeremy Burrows (MMV) for critical reading of the manuscript.

References

- of outstanding interest
- of special interest

1. Trouiller P, Olliori P, Torreele E, Orbinski J, Laing R, Ford N: **Drug development for neglected diseases: A deficient market and a public-health policy failure.** *Lancet* (2002) **359**(9324):2188-2194.
2. Pink R, Hudson A, Mouriès MA, Bendig M: **Opportunities and challenges in antiparasitic drug discovery.** *Nat Rev Drug Discov* (2005) **4**(9):727-740.
3. Chatelain E, Ioset J-R: **Drug discovery and development for neglected diseases: The DNDi model.** In: *Preclinical Drug Development: Genesis 2009.* University of Greenwich, London, UK (2009):52-61.
4. Nwaka S, Hudson A: **Innovative lead discovery strategies for tropical diseases.** *Nat Rev Drug Discov* (2006) **5**(11):941-955.
 - Provides a comprehensive description of the WHO/TDR, including the partners involved in the network, and the need for data management. This paper also describes challenges related to the diversity of partner cultures and processes, as well as the desire for IP protection.
5. Ribeiro I, Sevcik A-M, Alves F, Diap G, Don R, Harhay MO, Chang S, Pecoul B: **New, improved treatments for Chagas disease: From the R&D pipeline to the patients.** *PLoS Negl Trop Dis* (2009) **3**(7):e484.
6. **Diseases covered by Neglected Tropical Diseases department:** WHO, Geneva, Switzerland (2010). www.who.int/neglected_diseases/diseases/en/
7. Nwaka S, Ridley RG: **Virtual drug discovery and development for neglected diseases through public-private partnerships.** *Nat Rev Drug Discov* (2003) **2**(11):919-928.
8. Croft SL: **Public-private partnership: From there to here.** *Trans R Soc Trop Med Hyg* (2005) **99**(Suppl 1):S9-14.
9. Don R, Chatelain E: **Drug Discovery for neglected diseases: View of a public-private partnership.** In: *Antiparasitic and Antibacterial Drug Discovery: From molecular targets to drug candidates.* Selzer PM (Ed) Wiley-Blackwell, Hoboken, NJ, USA (2009):33-43.
 - Provides a description of the DNDi, including the management of the initiative's drug-discovery and development pipeline.
10. Bathurst I, Hentschel C: **Medicines for malaria venture: Sustaining antimalarial drug development.** *Trends Parasitol* (2006) **22**(7):301-307.
 - Provides a description of the MMV, including the initiative's large network of partners and the management of their project portfolio.
11. **TB Alliance: Global Alliance for TB Drug Development:** TB Alliance, New York, NY, USA (2010). www.tballiance.org
12. **Institute for OneWorld Health: Fulfilling the promises of medicine for the developing world:** Institute for OneWorld Health, San Francisco, CA, USA (2010). www.oneworldhealth.org/
13. **Accelerating malaria vaccine development:** Malaria Vaccine Initiative, Bethesda, MD, USA (2010). www.malariavaccine.org/
14. **FIND: Foundation for Innovative New Diagnostics:** Foundation for Innovative New Diagnostics, Geneva, Switzerland (2010). www.finddiagnostics.org/
15. Premji Z: **Coartem: The journey to the clinic.** *Malaria J* (2009) **8**(Suppl 1):S3.
16. Herring P: **Experiments in social responsibility.** *Nature* (2006) **439**(7074):267-268.
17. Kettler HE, Marjanovic S: **Engaging biotechnology companies in the development of innovative solutions for diseases of poverty.** *Nat Rev Drug Discov* (2004) **3**(2):171-176.
18. Vennerstrom JL, Arbe-Barnes S, Brun R, Charman SA, Chiu FCK, Chollet J, Dong Y, Dorn A, Hunziker D, Matile H, McIntosh K et al: **Identification of an antimalarial synthetic trioxolane drug development candidate.** *Nature* (2004) **430**(7002):900-904.
19. Searls DB: **Data integration: Challenges for drug discovery.** *Nat Rev Drug Discov* (2005) **4**(1):45-58.
20. Matter A, Keller TH: **Impact of non-profit organizations on drug discovery: Opportunities, gaps, solutions.** *Drug Discov Today* (2008) **13**(7-8):347-352.
21. Burrows J: **The state of the art in anti-malarial drug discovery and development.** *Current Topics in Medicinal Chemistry:* in press.
22. Hopkins AL, Witty MJ, Nwaka S: **Mission possible.** *Nature* (2007) **449**(7159):166-169.
23. Renslo AR, McKerrow JH: **Drug discovery and development for neglected parasitic diseases.** *Nat Chem Biol* (2006) **2**(12):701-710.
24. Wells TNC: **New paradigms for drug discovery from parasitology.** In: *Preclinical Drug Development: Genesis 2009.* University of Greenwich, London, UK (2009):39-45.
25. Herring PL: **Financing R&D for neglected diseases.** *Nat Rev Drug Discov* (2009) **8**(2):91.
26. Herring P: **Patent sense.** *Nature* (2007) **449**(7159):174-175.
27. Callan B, Gillespie I: **The path to new medicines.** *Nature* (2007) **449**(7159):164-165.
28. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT et al: **Genome sequence of the human malaria parasite *falciparum*.** *Nature* (2002) **419**(6906):498-511.
29. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renaud H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U et al: **The genome of the African trypanosome *Trypanosoma brucei*.** *Science* (2005) **309**(5733):416-422.

30. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream M-A, Adlem E, Aert R, Anupama A et al: **The genome of the kinetoplastid parasite, *Leishmania major***. *Science* (2005) **309**(5733):436-442.
31. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran A-N, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ et al: **The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease**. *Science* (2005) **309**(5733):409-415.
32. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S, Buck GA et al: **Complete genome sequence of the apicomplexan, *Cryptosporidium parvum***. *Science* (2004) **304**(5669):441-445.
33. Comini MA, Guerrero SA, Haile S, Menge U, Lünsdorf H, Flohé L: **Validation of *Trypanosoma brucei* trypanothione synthetase as drug target**. *Free Radic Biol Med* (2004) **36**(10):1289-1302.
34. Aguero F, Al-Lazikani B, Aslett M, Berriman M, Buckner FS, Campbell RK, Carmona S, Carruthers IM, Chan AWE, Chen F, Crowther GJ et al: **Genomic-scale prioritization of drug targets: The TDR targets database**. *Nat Rev Drug Discov* (2008) **7**(11):900-907.
- Describes the success of the TDR target genomics database and the results of a collaboration by geographically dispersed scientists.
35. **The TDR Targets Database: Identification and ranking of targets against neglected tropical diseases:** The TDR Drug Targets Network (2010). www.tdrtargets.org/
36. Aurrecochea C, Brestelli J, Brunk BP, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F et al: **EuPathDB: A portal to eukaryotic pathogen databases**. *Nucleic Acids Res* (2010) **38**(Database issue):D415-D419.
37. **TriTryp Database:** EuPathDB (2010). TriTrypDB.org
38. Desai PV, Patny A, Sabnis Y, Tekwani B, Gut J, Rosenthal P, Srivastava A, Avery M: **Identification of novel parasitic cysteine protease inhibitors using virtual screening. 1. The chembridge database**. *J Med Chem* (2004) **47**(26):6609-6615.
39. Desai PV, Patny A, Gut J, Rosenthal PJ, Tekwani B, Srivastava A, Avery M: **Identification of novel parasitic cysteine protease inhibitors by use of virtual screening. 2. The available chemical directory**. *J Med Chem* (2006) **49**(5):1576-1584.
40. Mehta SR, Huang R, Yang M, Zhang X-Q, Kolli B, Chang K-P, Hoffman RM, Goto Y, Badaro R, Schooley RT: **Real-time *in vivo* green fluorescent protein imaging of a murine leishmaniasis model as a new tool for leishmania vaccine and drug discovery**. *Clin Vaccine Immunol* (2008) **15**(12):1764-1770.
41. Williams AJ: **A perspective of publicly accessible/open-access chemistry databases**. *Drug Discov Today* (2008) **13**(11-12):495-501.
42. **Low-Hanging-Fruit: An antiparasitic drug database:** University of California, San Francisco, CA, USA (2010). www.sandler.ucsf.edu/lhf/
43. Ertl P, Jelfs S: **Designing drugs on the Internet? Free web tools and services supporting medicinal chemistry**. *Curr Top Med Chem* (2007) **7**(15):1491-1501.
44. Maurer SM, Rai A, Sali A: **Finding cures for tropical diseases: Is open source an answer?** *PLoS Med* (2004) **1**(3):e56.
45. Pécol B: **Neglected diseases: Wake up!** *Medicus Mundi Schweiz Bulletin* (2007) **103**:36-41.
46. **Hit Explorer Operating System (HEOS):** SCYNEXIS Inc, Research Triangle Park, NC, USA (2010). www.heos.com
47. Hohman M, Gregory K, Chibale K, Smith PJ, Ekins S, Bunin B: **Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery**. *Drug Discov Today* (2009) **14**(5-6):261-270.
- Describes a SaaS platform for drug discovery.
48. Nwaka S, Ramirez B, Brun R, Maes L, Douglas F, Ridley R: **Advancing drug innovation for neglected diseases – Criteria for lead progression**. *PLoS Negl Trop Dis* (2009) **3**(8):e440.
- Presents the results of research conducted by the WHO/TDR to standardize the terms and processes within its network of partners.
49. **Lead discovery for infectious tropical diseases – 2008 annual report**. Special Programme for Research & Training in Tropical Diseases, Geneva, Switzerland (2009). apps.who.int/tdr/svc/publications/about-tdr/annual-reports/lead-discovery-for-drugs-2008
50. Jacobs R, Nare B, Wring S, Bacchi C, Brun R, Plattner J, Beaudet B, Bowling T, Chen D, Freund Y, Gaukel E et al: **Efficacy and pharmacokinetics of SCYX-7158 (AN 5568): A novel and potent oxaborole-6-carboxamide selected as a pre-clinical candidate for once-daily oral treatment for stage 2 human African trypanosomiasis**. *Meeting of the American Society of Tropical Medicine and Hygiene*, Washington, DC, USA (2009) **58**:Abs 2980.
51. **ActivityBase Suite:** ID Business Solutions Ltd, Guildford, UK (2010). www.idbs.com/activitybase/
52. Hassan M, Brown R, Varma-O'Brien S, Rogers D: **Cheminformatics analysis and learning in a data pipelining environment**. *Mol Divers* (2006) **10**(3):283-299.
- Provides an example of an analysis of scientific data that can be conducted with cheminformatics tools such as Pipeline Pilot.
53. **TIBCO Spotfire:** Tibco Software Inc, Somerville, MA, USA (2010). spotfire.tibco.com/
54. **The R project for statistical computing:** The R Foundation, Vienna, Austria (2009). www.R-project.org
55. Barber CG, Haque N, Gardner B: **'One point' – Combining OneNote and SharePoint to facilitate knowledge transfer**. *Drug Discov Today* (2009) **14**(17-18):845-850.
- Provides a good example of how software such as Microsoft Sharepoint can aid drug-discovery collaborations.
56. Williams AJ: **Internet-based tools for communication and collaboration in chemistry**. *Drug Discov Today* (2008) **13**(11-12):502-506.
57. Java A, Song X, Finin T, Tseng B: **Why we twitter: Understanding microblogging usage and communities**. *Joint 9th WebKDD and 1st SNA-KDD 2007 workshop: Web mining and social network analysis*, San Jose, CA, USA (2007):Abs ID21.
- Provides a good illustration of the new type of communication software proposed by the consumer market.
58. Acquisti A, Gross R: **Imagined communities awareness, information sharing, and privacy on the facebook**. *Workshop on Privacy Enhancing Technologies*, Cambridge, UK (2006) **6**:36-58.
59. Debatin B, Lovejoy JP, Horn A-K, Hughes BN: **Facebook and online privacy: Attitudes, behaviors, and unintended consequences**. *J Comput-Mediat Commun* (2009) **15**(1):83-108.
60. Ellison NB, Steinfield C, Lampe C: **The benefits of facebook "friends": Social capital and college students' use of online social network sites**. *J Comput-Mediat Commun* (2007) **12**(4):1143-1168.
61. Bailey DS, Zanders ED: **Drug discovery in the era of Facebook new tools for scientific networking**. *Drug Discov Today* (2008) **13**(19-20):863-868.
- Provides a review of existing Internet-based drug-discovery communities.
62. **Orchid neglected diseases drug development community:** LinkedIn Corp, Mountain View, CA, USA. www.linkedin.com/groups?home=&gid=1783204&trk=anet_ug_hm
63. Warr W: **A grand vision for configurable science and minimizing the loss model**. *J Comput Aided Mol Des* (2008) **22**(8):587-591.
64. **ChemBioConnect: Collaborative visual chemistry research tool:** Imaginatik plc, Boston, MA, USA (2010). www.imaginatik.com/webdoc_prod_chembioconnect
65. Symyx Solutions Inc: **Symyx announces ChemMobi application now available on app store**. *Press Release* (2009): July 23. www.symyx.com/company/news/press_releases/2009/pr_chemmobi_app_store.jsp

66. **Welcome to ChemJuice:** ID Business Solutions Ltd, Guildford, UK (2010). www.idbs.com/ChemJuice/
67. **Salesforce Chatter:** Salesforce.com Inc, San Francisco, CA, USA (2010). www.salesforce.com/chatter
68. Skeels MM, Grudin J: **When social networks cross boundaries: A case study of workplace use of Facebook and LinkedIn.** In: *Proceedings of the ACM 2009 International Conference on Supporting Group Work*. Association for Computing Machinery, New York, NY, USA (2009):95-104.
 - Provides an analysis of the strengths and weaknesses of the use of social networks for professional applications.
69. **Strategic and business plan for the African Network for Drug and Diagnostics Innovation (ANDI):** WHO, Geneva, Switzerland (2009). meeting.tropika.net/andi2009/files/2009/10/sbp-final_web.pdf
 - Provides a strategic business plan for the creation of a PPP, including a description of the required information technology and data management solutions.
70. Pécoul B: **Interview with Dr Bernard Pécoul, Executive Director of the Drugs for Neglected Diseases initiative.** *Future Medicinal Chemistry* (2009) **1**(2):243-248.