
Screening and detecting ligand binding to biological macromolecules by circular dichroism

Abstract

The subject invention presents a method for screening potential ligands of biological macromolecules by synchrotron radiation circular dichroism (SRCD). Protein samples in solution will exhibit conformational changes when bound to a ligand significantly different from protein samples which are unbound. The subject invention is intended for the discovery of novel ligands which bind to HIV-1 and HIV-2 gp120, gp41, reverse transcriptase, and protease; however it can be used for analysis of any biological macromolecule. The subject invention includes software for automatically detecting conformational changes between SRCD spectra and an automated sample changer to increase throughput screening of samples.

Inventors: **Engel; Matthew A.** (Ronkonkoma, NY), **Allaire; Marc** (Upton, NY)
Assignee: **Brookhaven Science Associates, LLC** (Upton, NY)
Filed: **December 12, 2006**

U.S. Patent Documents

7130747	October 2006	Von Dreele, R. et al.
6843262	January 2005	Ismagilov, R. et al.
6818740	November 2004	Eckert, et al.
6821744	November 2004	Von der Eltz, H., et al.
6596497	July 2003	Jiang, S. et al.
6506554	January 2003	Chan, D. et al.
5780242	July 1998	Nickel, A.

Other References

Brookhaven National Laboratory, N. S. L. S. (2004) VUV Storage Ring Parameters as of January 2004.
Canaan, S., Zadori, Z., Ghomashchi, F., Bollinger, J., Sadilek, M., Moreau, M. E., Tijssen, P., and Gelb, M. H. (2004) Interfacial enzymology of parvovirus phospholipases A2, *The Journal of biological chemistry* 279, 14502-14508.
Coffin, John and S. H., Harold Varmus (1997) *Retroviruses*, Cold Spring Harbor Laboratory Press.

Claims

The invention claimed is:

1. A method for screening and detecting ligand binding to macromolecules in a sample solution using ultraviolet (UV) radiation circular dichroism.
2. A method of claim 1, in which UV radiation for circular dichroism is produced by a dedicated synchrotron storage ring emitting vacuum ultraviolet (VUV) synchrotron radiation; herein known as synchrotron radiation circular dichroism (SRCD).
3. A method of claim 1, in which radiation for circular dichroism is produced by commercially available benchtop circular dichroism instruments; herein known as circular dichroism (CD).
4. A sample solution according to claim 1, wherein said sample solution contains the biological macromolecule or macromolecules of interest.
5. A sample solution according to claim 1, wherein said sample solution contains the biological macromolecule or macromolecules of interest and a known ligand with potential binding capacity to that macromolecule.
6. A method of claim 1, in which sample solutions from claim 4 and sample solutions from claim 5 can be differentiated based on conformational properties.
7. A method of claim 1 where fluctuations between circular dichroism spectra from different sample solutions are automatically identified by statistical software analysis.
8. A biological macromolecule according to claim 4, wherein said macromolecules are: proteins, nucleic acids, protein-nucleic acid complexes, glycoproteins, proteoglycans, saccharides, glycolipids, lipids, and lipoproteins.
9. A protein according to claim 8, wherein said protein can be a biologically active or inactive polymer chain composed of amino acids linked by peptide bonds. The protein may be a capsid protein, envelope protein, membrane-associated receptor, nuclear receptor, cytosolic receptor, enzyme, kinase, transcription factor, tumor suppressor, growth factor, and cytoskeletal protein filament.
10. A protein according to claim 8, wherein said protein is HIV-1 and HIV-2 is reverse transcriptase and protease. Said protein may contain the native amino acid sequences from any subtypes found in the National Center for Biotechnology Repository or be modified through natural deglycosylation, sulfhydryl crosslinking, truncation, point mutation and other covalent modification.
11. A glycoprotein according to claim 9, wherein said glycoprotein is gp120 and gp41 from HIV-1 and HIV-2. Said glycoprotein may contain the native amino acid sequences from any subtypes found in the National Center for Biotechnology Repository or be modified through deglycosylation, sulfhydryl crosslinking, truncation, point mutation and other covalent modification.

12. A ligand according to claim 5, wherein said ligands can be proteins, peptides, nucleic acids, metals, ions, polymers, antibiotics, and small organic molecules.

13. A protein according to claim 12, wherein said protein ligands are drugs, antibodies, neurotransmitters, and hormones.

12. A ligand according to claim 12, wherein said ligand will be mixed with potential binding partners at low concentrations in the range of micromolar to nanomolar dose formulations.

14. A nucleic acid according to claim 12, wherein said nucleic acids are genomic DNA, genomic RNA, small interfering RNA, and messenger RNA from prokaryotes, eukaryotes, and viruses.

15. A small organic molecule according to claim 12, wherein said small organic molecule consists of a scaffold with variable reactive functional groups covalently linked at specific points. Its molecular weight will typically be under 500 Da to insure solubility. The small organic molecule may have many distinct analogues with different bioactivities. The small organic molecule may have pro-chiral stereoisomers with different bioactivities. The organic molecule may also be a solvent or denaturant such as ethanol, urea, and guanadinium hydrochloride.

16. A method of claim 2, in which SRCD data collection occurs in a sample chamber with an automatic sample changer. The sample chamber is purged with gaseous nitrogen or other gases to displace atmospheric oxygen prior to and during data collection.

17. A method of claim 2, in which SRCD data is collected by an automatic sample changer which is a mechanical device consisting of a two motors which rotate sample cells post data collection. Sample cells are made of quartz or calcium fluoride. Each sample cell contains 3 repositories for various sample solutions. Each sample cell will be rotated in place 120° when data collection is complete initiating SRCD data collection from the neighboring repository. Following data collection from all cell repositories in a single sample cell, the cell changer rotates 60° to begin collection from the neighboring cell.

18. A method of claim 2, where the ambient temperature of the sample chamber may be controlled, specifically ranging from 0°C to 120°C.

19. A method of claim 17, in which the sample cell is composed of UV-transparent calcium-flouride material.

BACKGROUND OF THE INVENTION

The subject invention is to be utilized for the discovery of novel ligands which interact specifically with biological macromolecules in solution. The method of discovery relies on high energy ultraviolet radiation its application to a sample in solution

producing a spectral image. Spectra distinguish binding from non-binding ligands on the atomic scale by illustrating a change in sample structure, if present. Rearrangement or interference with macromolecular structure often results in a change or loss of function by that macromolecule. Spectra are produced by synchrotron radiation circular dichroism (SRCD) or circular dichroism (CD) and can reveal alterations in macromolecular structure. The macromolecules will often be proteins, but are not limited in scope.

Development of a pharmaceutical product is a laborious and lengthy process. Advancements in gene sequencing and recombinant protein expression have made hundreds of new drug targets rapidly available to researchers and pharmaceutical companies for experimentation. To streamline this process, an expressed protein may be produced in large quantities and distributed to a prefabricated sample cell. The initial 'hit' of a lead compound with its intended target requires the screening hundred's of thousands of samples. These samples contain small quantities of biological macromolecules and potential ligands in a mixture. The potential ligands are often present in micromolar quantities to insure specificity. The sample cells each contain one particular condition selected from a combinatorial library of potential ligands with known chemical structure and composition.

Many pathological proteins of interest originate from neoplastic growths, prokaryotic bacteria or viruses. Traditional biochemical assays have relied on enzymatic kinetics to quantify the degree of inhibition based on the accumulation of products or consumption of substrate. However, such methods require the development of unique assays for each target, which is impossible to create for targets with unknown function. Therefore it is desirable to have a method for drug discovery which does not require previous knowledge of the protein's function. The only requirements include a gene and complimentary suite of ligands with potential binding properties heralding the era of chemical genetics. It is predicted that the present invention will be useful for discovering previously unidentified ligand-receptor interactions.

The subject invention is capable of imaging any macromolecule in solution, however, the first embodiment specifically targets Human Immunodeficiency Virus (HIV) clades 1 and 2. Retroviruses, such as HIV, contain genomic RNA but are replicated by an integrated DNA complex. The retrovirus ~9kb genome contains a promoter at the 3' end which drives the expression of reverse transcriptase directly from this material. The viral genes *gag*, *pol*, *pro*, and *env* encode the major retroviral proteins. *Gag* encodes a full-length protein which is proteolytically cleaved into mature protein (MA), capsid (CA), and the nucleocapsid (NC). *Pol* encodes for reverse transcriptase and integrase enzymes. An incomplete list of claimed macromolecules imaged by the present invention include: glycoprotein 120 (gp120), glycoprotein 41 (gp41), HIV reverse transcriptase, HIV integrase, and HIV protease present in all viral subspecies.

HIV-1 infection is initiated by high-affinity binding of envelope proteins (ENV) to the human CD4 receptor. A pivotal coreceptor in the immune-pathway, CD4 stimulates association between T-cell receptors and major histocompatibility complex II which presents antigens on the surface of infected cells. The CD4 extracellular domain exhibits dimeric properties which may behave in a hinge-like manner important for immune recognition and HIV binding. CD4⁺ lymphocytes are efficiently destroyed by HIV-1 and HIV-2 leading to acquired immunodeficiency syndrome (AIDS). HIV entry is dependent on interactions between the CD4 glycoprotein and gp120, an exterior viral

envelope glycoprotein and immunological target during infection. High resolution crystal structures of this antigen-antibody-receptor complex reveals seven disulphide bridges conserved among HIV strains. Gp120 is folded into two parallel domains (inner/outer) and composed of five variable regions (V1-V5) secured to the viral membrane by gp41, a transmembrane envelope glycoprotein. The peripheral region of this complex forms a spike, composed primarily of gp120 affixed by non-covalent interactions to gp41. This complex is utilized in direct membrane penetration and elicits high titers of ineffectual antibodies throughout infection. Upon binding to CD4, a conformational change arises in gp120 exposing sites for further coreceptor attachment.

The subject invention will exclude any drugs discovered that inhibit contact between the gp41 N36 and C34 peptide trimers as protected under U.S. Patent No. 6,506,554. Many small molecules, such as U.S. Patent No. 6,818,740, exist which claim to prevent HIV membrane fusion by inhibiting the binding of gp41 N and C terminal coiled-helices. Other ligands which bind to macromolecules such as HIV protease and reverse transcriptase consist of a drug scaffold and an array of functional groups conjugated at specific points.

Synchrotron radiation circular dichroism (SRCD) is an emerging technique available to biologists and chemists invested in structural genomics, chemical biology, pharmacology, drug discovery and structural biology. SRCD is comparable to circular dichroism (CD) data taken on bench-top equipment; however SRCD offers improved signal-to-noise and lower wavelength data unobtainable on bench-top equipment. Spectra above 190 nm will often be equivalent. Therefore, the present invention encompasses a method for screening macromolecules with one compound or library of compounds by SRCD and CD.

The strength of CD relies on fundamental optical characteristics of chiral macromolecules. These chiroptical properties give rise to a signature spectrum particular to one backbone conformation. Fluctuations in backbone conformation result in convolutions of spectra. Polypeptide chromophores exhibit electronic transitions within a molecular orbital as a function of secondary structure which generate peaks at particular wavelengths and compose said spectrum. The macromolecule may be produced by recombinant protein expression or isolated from tissues as per U.S. Patent No. 5,780,242 by Nickel, A. Nickel's invention describes the isolation of ion channels from patient's tissue samples in order to subject them to CD. This method claims to distinguish healthy from diseased cells for diagnosis of an ion channel disease.

U.S. Patent No. 6,821,744 to von der Eltz et al. describes a chemiluminescent assay in which HIV protease forms a detectable complex when bound to an inhibitor. This method excels at quantifying the inhibition of HIV protease but is incapable of probing the structure or function of any other macromolecules. U.S. Patent No. 6,972,126 to Allaway et al. describes a method for detecting HIV envelope fusion to CD4 by resonance energy-transfer. U.S. Patent No. 6,596,497 to Jiang et al. describes a screening method used to quantify solid phase monoclonal antibody binding in to the HIV-1 gp41 core for inhibitor identification. Each of these inventions share the same weakness, which is the inability to vertically transfer their application to new targets.

U.S. Patent No. 7,130,747 to Von Dreele et al., describes a method for x-ray diffraction of microcrystals in solution in order to detect ligand binding. The invention does not require single-crystals, and instead only requires a slurry of microcrystalline

material for diffraction. Conditions favorable for microcrystal growth are broader compared to conditions for single crystal growth making microcrystalline conditions easier to identify. X-ray powder diffraction of polycrystalline macromolecular material will rapidly produce images on the millisecond timescale amenable to high-throughput screening. The powder diffraction method may be more sensitive than the present invention because microcrystals contain a regular atomic lattice with ligand-bound perturbations reflected in the diffraction. However, SRCD and CD eliminate the need for any crystal growth and only require a macromolecule in solution bypassing the limitation of the aforementioned method.

Both SRCD and CD can identify substrate binding partners between any macromolecule and its explicit ligand by detecting conformational changes that the macromolecule exhibits under UV radiation. Additionally, the binding partner does not have to interact with the active site of a macromolecule to change its conformation. Ligands which do not bind to an active site but do bind in another region may not inhibit enzymatic activity escaping detection by enzymatic assay. However, said ligand may still have a potent affect in vitro and in vivo. For example, the capsid of erythrovirus B19 contains an N-terminal phospholipase A₂ (PLA₂) domain responsible for the hydrolysis of phospholipids during infection. Ligands which bind to this protein and alter its conformation may not bind to the PLA₂ active site, but instead adhere to proximal loops. While PLA₂ retains its functionality on the capsid surface, the ligand bound state may be so malformed as to prevent viral infection.

The subject invention is therefore embodied by two related methods for detecting conformational changes in a macromolecule when exposed to potential ligands. These methods consist of SRCD and CD solution screening. In the subject invention, solutions under investigation will consist of a biological macromolecules and potential ligands or biological macromolecules alone for comparison's sake. SRCD will specifically target HIV-1 and HIV-2 gp120, gp41, reverse transcriptase, and protease in order to discover novel ligand binding sites, observe modalities of conformational change, and uncover previously unknown ligands that bind to these macromolecules, not necessarily at the active site.

SUMMARY OF THE INVENTION

The subject invention provides a method for determining whether an agent is capable of specifically altering the conformation of macromolecules, such as proteins and glycoproteins, in the solution state when bound to that agent. The method of the subject invention consists of synchrotron radiation circular dichroism and circular dichroism during which the mixture, containing sample and agent, is placed in the path of a circularly polarized beam of light.

The subject invention serves to detect interactions between ligands and macromolecules on the atomic level. These interactions may be covalent, electrostatic (ionic, hydrogen bonded), or hydrophobic in nature and modulated by van der Waals, Lennard-Jones, and solvation free energies which in sum, determine the structure of a biological macromolecule at a given instant in time.

In the preferred embodiment, the circularly polarized light will be produced by high energy synchrotron radiation operating at energies greater or equal to 600 MeV

herein referred to as synchrotron radiation circular dichroism (SRCDD). In the preferred embodiment, the nominal operating energy will exceed 800 MeV providing the most accurate data for the collection of sensitive measurements and discernment between SRCDD spectra. The preferred current of the vacuum ultraviolet (VUV) ring will meet or exceed 1 Ampere at 700 MeV. These characteristics provide longer beam life extending collection time and maximizing high energy data collection.

In another embodiment, the circularly polarized light will be produced by an air-cooled or water-cooled Xenon lamp.

In the preferred embodiment, the subject invention will provide circular polarization measurements from 160 nm to 320 nm collected at 1 nm intervals. The present invention consists of an:

- a) incident photon beam (I_0) which passes through a
- b) circular polarizer transforming linearly polarized light into circularly polarized light which transverses the sample chamber.
- c) photoelastic modulator to control the frequency of polarization
- d) sample chamber holding the macromolecular target and ligand solution and
- e) photomultiplier tube (PMT) for the detection of photons from beam I_0 .
- f) a PC connected to a lock-in amplifier that read the CD signal from the PMT
- i) software on the PC for rapid collection and comparison of CD spectra.

The sample chamber will contain an apparatus used for automatically changing samples following data collection from each sample (figure 1). This apparatus will be called an automatic sample changer and contain multiple sample cells which enclose the various solutions under investigation.

In the primary aspect of the preferred embodiment, macromolecules in solution are composed of recombinant HIV-1 and HIV-2 glycoproteins and enzymes expressed in bacteria or mammalian cells. This technique will allow the experimenter to control biochemical attributes of the particular domain under investigation and facilitate downstream purification. Recombinantly engineered proteins may exhibit fusion partners, artificial covalent modifications, protease cleavage sites, or point mutations in order to facilitate purification or investigate the effects of these modifications on ligand binding and secondary structure.

In the secondary aspect of the preferred embodiment, macromolecules in solution are composed of glycoproteins, proteins or nucleic acids isolated directly from pathological virions, infectious bacteria, or diseased tissue.

In the preferred embodiment, the automatic sample changer will streamline data collection by SRCDD and CD. A beam sensor upstream from the sample changer will prevent data collection from occurring when there is no beam. For example, during fills of the synchrotron UV ring which occur every several hours. The automatic sample changer will contain cells which must be prefilled with sample solutions prior to data collection.

DETAILED DESCRIPTION OF THE INVENTION

In the preferred embodiment ligands will be obtained from both public and private drug libraries. In one aspect of the preferred embodiment, ligands are obtained from public resource facilities. Public databases offer free or discounted pharmaceutical

compounds to academic research institutions. The Developmental Therapeutics Program (DTP) through the National Cancer Institute (NCI) and the National Institutes of Health (NIH) offers many publicly available libraries. Ligands in these libraries may or may not be targeted to proteins of a particular function.

In the first aspect, DTP offers an archive of more than 140,000 non-discreet synthetic and naturally occurring compounds. Small numbers of specific agents or large quantities of multiwell plates can be obtained through this mechanism. Plated ligands include diversity, mechanistic, challenge, and natural product sets which all contain potential therapeutic moieties and tight-binding products of biological macromolecules. The Natural Product Repository from NCI contains over 50,000 plant samples from Africa, Central and South America, and Southeast Asia. Compounds are provided in a 1:1 mixture of dichloromethane and methanol to water. The reagents are stored at -20°C. Extracts are provided in vials or 96 well plates for distribution through the open repository program. The Biological Products Laboratory Repository from NCI contains monoclonal antibodies targeted against HIV which are free after approval by the FCRDC AIDS reagent approval committee.

Also in the first aspect, the National Institute of Allergy and Infectious Diseases (NIAID) provides a resource guide for the development of AIDS therapeutics. While the NIAID does not provide compounds for biochemical assays, it does provide a searchable, computerized database linking all known drugs with their chemical structures, preclinical antiviral and antimicrobial efficacy data. Upon request, information regarding the patent status of drugs in the database can be obtained from the DIAIDS staff, making this resource indispensable to the inventors.

In the second aspect of the preferred embodiment, ligands are obtained from private research facilities. This includes any vendor listed in the ZINC, non-commercial database which provides links to 1000's of private vendors and their available ligands.

In the preferred embodiment, proteins and genes to be expressed be obtained from public research universities and government agencies. For example, in one aspect the NIH AIDS Research and Reference Reagent Program will provide HIV Env-1 clones from subtypes B and C.

The preferred embodiment consists of synchrotron radiation circular dichroism (SRCD). SRCD is a developing technique with significant advantages over traditional circular dichroism (CD) spectroscopy. Primarily, the intense ultraviolet (UV) light produced by synchrotron radiation facilities provides data at lower wavelengths unobtainable by conventional measures. Compiling measurements over a broader spectrum in the low-wavelength far UV results in a more complete waveform with more structural data, possibly revealing conformation changes that were not visible before. This acquisition is possible due to the concentrated flux of photons provided by the synchrotron in the vacuum UV (VUV) range. The strength of CD relies on the fundamental chiroptical properties of proteins in which backbone fluctuations result in convolutions of spectra. Polypeptide chromophores exhibit electronic transitions within a molecular orbital as a function of secondary structure which generate peaks at particular wavelengths.

Currently, there is a substantial variety of analytical techniques capable of interpreting CD data. These algorithms, designed independently of each other, produce inter and intra-variable results depending on the reference set used, lowest wavelength

collected, and accuracy of pre-specified measurements such as protein concentration and pathlength. Aside from heuristic approaches, statistical analysis of raw data provides a complementary description of fluctuations in secondary structure between samples. This is particularly useful for inspecting accuracy, assessing error and importantly, detecting conformational change. In essence, data in the lowest wavelength range is subject to exclusion when reproducibility dissipates. Here we observe minimal deviation between spectra down to 173 nanometers (nm) at which circular dichroism data became attenuated by the adsorption of water, buffer and the sample. It has been found that UV transparent cells may reduce this interference.

The impact of low wavelength data is significant. Spectra below 178 nm are information rich and exhibit spectral formations embodying structural characteristics undetectable at wavelengths over 190 nm. Nearly twice as much structural information can be derived from data taken in the VUV range (between 178 nm and 168) compared to spectra truncated at 190 nm. The presence of low wavelength data allows experimenters to examine fluctuations in secondary structural content with greater accuracy. These results are constrained by the reference sets available during CD data analysis. Currently, SP175 from Wallace et al. is the only reference database capable of analyzing spectra containing low wavelength data and is truncated at 175 nm.

In the preferred embodiment, the PC will contain software capable of quickly distinguishing between spectra of different conformations using a neural network. Current in-house software is capable of displaying collective SRCD spectra during data accumulation. Current publicly available academic software is capable of displaying collective SRCD spectra during data accumulation, subtracting spectra from each other, multiplying spectra by constant and overlaying stored spectra for manual analysis.

The subject invention consists of software capable of analyzing SRCD spectra a) at particular wavelengths central to structural composition b) by integrating the area between different spectra c) analyzing this area, or RMSD, for statistical significance d) quickly providing a secondary structural analysis using a neural net, such as K2D, but instead using data down to 178 nm.

In the preferred embodiment, the automated sample changer consists of a circular metal surface capable of being rotated at regular intervals upon receiving input from a beam sensor or timer attached to a PC. The circular metal surface is thick enough to rest up to ten sample cells which are only several millimeters in width. When mounted vertically, the circular metal surface will keep the sample cells in place with rubber flaps as to prevent them from falling out and being scratched or damaged. Upon completion of data collection from a single reservoir in the sample cell, a small mechanical belt will rotate the sample cell 120° aligning the synchrotron beam with the neighboring reservoir. Upon completion of data collection from all three reservoirs, a center motor mount will rotate the circular metal surface 36° aligning the first reservoir from the neighboring sample cell with the synchrotron beam. This is repeated until data collection is complete from all sample reservoirs in all sample cells.

In the preferred embodiment the sample cell is composed of calcium fluoride. In a second embodiment the sample cell is composed of quartz. Current cells designed by Hellma (DE), can be bought and machined from either material at various pathlengths up to 4 microns in width. However, the subject invention improves upon these cells by consisting of three sample reservoirs instead of just one. In each embodiment, the cells

will be airtight, as per cells fabricated by Hellma, preventing the evaporation of solvent from the mixture during data collection.