Application of Computer Algorithms for Their Use in Chemistry for Elucidation of Protein Ligand Docking

Sarla Namdeo Kalyankar, Namdeo Venkatrao Kalyankar

***Abstract*-- Here the method to address the docking problem by algorithmic concepts is given. These algorithms are developed based on the problem arises from the structures stored in the Protein Data Bank (PDB) [1]. The Atomic coordinates of the complex associates method for predicting the structure of protein Ligand and protein-protein complexes was employed.**

The algorithm consists of three independent substeps:

1. **Classification of molecular surfaces**
2. **Determination of complex structures by means of complementary surface properties**
3. **Summarizing similar complex structures to a middle structure using a Cluster algorithm**

***Index Terms*— Algorithm, Ligand, Geometric, Docking, Molecular**

I. Introduction

The subdivision of molecular surfaces into regions with similar topographic or physicochemical properties, hereinafter referred as domains designated as used. The resulting domains of the two complex partners become compared with the algorithm and suggestions for possible complex structures were made. These complex structures can be combined and can be processed in any order. So similar structures using a cluster algorithm to a middle complex structure is summarized. In addition, the complex structures with the downhill simplex method or with a optimized genetic algorithm. Through the downhill simplex algorithm, the ligand is transformed into the nearest one using rigid molecular structures with local minimum movement. This makes the Downhill Simplex algorithm suitable for pre-optimization for a variety of possible complex structures. The genetic algorithm, on the other hand, takes into account the flexibility of the ligand and is intended to precede all serve for the flexible optimization of the most energetically favorable complex structure.

II. Segmentation of Molecular Surfaces

The calculation of the Solvent Accessible Surface results according to the canonical curvatures three different types of surface areas: concave and convex areas and saddle areas [2]. Lin et al. [3] developed a procedure that each of these surface areas by a representative surface point (sparse Critical Point) and a mean normal vector. The resulting classification of the molecular surface has the disadvantage that only very small Surface areas are described and thus on each surface a large number arise from reference points. For further processing, the number of Reference points are further reduced by adding very close to each other areas one and for each complex partner only one particular type of Surface areas, e.g. concave areas of the receptor and convex areas of the receptor Ligands, used for the comparison [4].

Heiden et al. [5] developed based on the fuzzification of the STI, the electrostatic potential and lipophilicity molecular surface into larger, cohesive surface areas. The achievement for this can be done for all surface areas the affiliation to the corresponding structural variables calculated and adjacent points with similar properties to domains. However, in the algorithm proposed by Heiden, domains arise in their size and shapes are very different and therefore difficult to each other can be compared. Therefore, the following is a segmentation algorithm presented, which allows the expansion to larger surface areas and still generated easily for comparable surface domains. The algorithm is based on the principle of the growth of the domain starting from critical points up to a maximum size defined by the user. The critical Points are considered local maxima of belonging to a class of structural Variables determined. The domains then become the class with the largest Assigned relationship value of the critical point. The topography of the domains of the classes, crevice, saddle, burr and graft are considered. This is referred to as a relationship of a domain to a particular class in the Contrary to Heiden et al. [6] proposed segmentation of surface points are assigned to multiple domains. So it comes to one overlapping of the domains. The boundaries of the domains are not sharply defined but surface point has an affiliation with each domain. So arises for each domain is a small core area with surface points with a high degree of relationship to the corresponding domain. To enlarge the domain the use of points correspondingly smaller affiliation (α -level amount with small value α) can be taken into account. The maximum size of the domain is finally reached, if no surface point more has a relationship above zero to the corresponding domain. For all classes of structural variables

The first step of the segmentation is the calculation of the

Manuscript received March 06, 2018; revised March 30, 2018.

Sarla Namdeo Kalyankar, Assistant Professor, Department of chemistry, Yeshwant Mahavidyalaya, Nanded, Maharashtra India phone: +91-2462-269600; fax: +91-2462-253726 (e-mail: drkalyankarsn@ yahoo.com)

Namdeo Venkatrao Kalyankar, Vice-Chancellor, Gondwana University, Gadchiroli, Maharashtra, India phone: +91-7132-222221; fax: +91-7132-223105 (e-mail: drkalyankarnv@yahoo.com)

ISBN: 978-988-14047-9-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

relationship of the classes of the structural variables of the corresponding property. As basis variables can be the shape index and the electrostatic potential as well as the lipophilicity was used. The individual classes of structural variables are consecutively segmented used by the surface points of increasing affiliation to the examined class of structural variables. The surface point with the largest affiliation- the central point is the center of the new domain selected. The similarity or dissimilarity of all surface points to the central point was calculated. With an area of one one-third of the maximum size of the domain around the central surface points, the one Similarity of the focus of these points and the central point was determined. This focus turns into a new domain center calculated. This results as the surface point of a projection of the center of gravity nearest to the surface. This ensures that the new domain center at the center of a local extreme of the property being studied located. Starting from this domain center, all neighboring points along the surface determined until the surface area defined by these points the maximum size of a domain have reached. For the determination of the neighboring points, the algorithm used to calculate the global curvatures (For each of these neighboring points, the shortest path along the Triangular edges determined to the center of the domain. A domain then becomes out of all Points whose similarity to the central point is greater than 0.5 (α -level with α = 0.5) and no surface point on their pathway to the domain center whose similarity is less than 0.5. Is the size of the domain smaller than one certain minimum size, the domain in the further bill is no longer considered. Each surface point of a domain is assigned a relationship value consisting of calculated from the distance to the center of the domain.

For all points of the domain results in a relationship greater than zero, for all others points equal to zero. The domains thus generated have an approximately circular shape. This makes the domains comparable. The different size of the domains on opposite parts of the two complex partners can be identified by the fuzzy definition of the domain boundary are matched. The comparison of a large domain only surface points with a large one domain relationship , whereas for small domains an α - Low-level domain relationship of small domain α is used. The segmentation algorithm then becomes the surface point with the next largest Belonging to the examined class, which has not yet been assigned to a domain of the same class is, continued. As central points for the domain of a certain class but only Surface points whose relationship in this class is greater than all others work other classes were functional. Otherwise, a domain would be defined by a surface point its properties have little or no relevance to the criteria assigned to the class of the domain do not match. If there is no more surface point that has a class affiliation above a certain minimum value, the next class becomes the structural one variable continued.

III. Comparison of the Domains of the Molecular Surfaces

The molecular surfaces of both complex partners described above by the method of segmentation based on the shape index, the electrostatic potential and lipophilicity were studied. By comparing the properties of the so generated domains can be used to identify suggestions for a possible complex structure. The Individual steps are described in as follow.

A. Determining the properties of a domain

The topographic or physicochemical properties of a domain are determined by the Mean value of the corresponding properties of the surface points defined. To get one in later algorithm to allow necessary adjustment of the size of the domains these averaged properties not only for the entire domain but also for calculated different α -levels of domain relationship. This is the Average of all surface points that belong to the corresponding α -level quantities were determined. In addition to each of these α -levels is the number the surface points contained and the size of the resulting surface certainly. The description of the location of the domain is made by a representative point and a normal vector. The representative point is the focus of the Surface points of a core region of the domain. The core area will be all Points with a domain affiliation greater than 0.9 are considered. The normal vector is calculated as the mean value of the normal vectors of the points of the core area.

B. Comparison of Domain Properties

For the formation of a stable complex with one complementarity of the molecular surfaces is necessary for the complex formation, this should be described by the existence of complementary surface domains. The algorithm presented here quantifies with the aid of a comparison of the Domain properties the complementarity of any possible combination of a domain of the receptor with a domain of the ligand and determines possible Bonding regions on the molecular surfaces. First, a negative image of the receptor is created by giving to each domain as an ideal complementary domain as A'. The properties of domain A' arise according to the following scheme: The size and the number of surface points of A' correspond to the original domain. This corresponds to a mapping of a convex domain to a concave domain with the same ratio of the two global curvatures and vice versa. A saddle area becomes a saddle area again assigned. To ensure equal absolute amounts of the curvatures of A' and A, the complementary domain receives the original value of the curvedness. The domains of the negative Image of the receptor classified according to electrostatic potential or lipophilicity were obtained, the negative value of the averaged electrostatic potential or the Original value of the lipophilicity of A. For further comparison, for all combinations of receptor domains A and Ligand domains B corresponding α -level levels determined, the sizes of which are most likely correspond. This allows for an adaptation of the domain size, and it can not only the properties averaged over the entire domain, but also one of Core area up to the size of the smaller domain going profile of the properties be compared.

The complementarity of the electrostatic potential and the lipophilicity cannot be attributed to the same way be calculated. While in the topography a complementary Surface can be described by the ideally complementary

ISBN: 978-988-14047-9-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

domain are in these Properties a variety of different definitions with ideal complementarity possible. Domains of the electrostatic potential must be different signs, domains of lipophilia have the same sign.. For the calculation of the complementarity of Domain combinations with the same sign of the electrostatic potential or different sign of lipophilicity becomes neutral on the class of the corresponding structural variables are used. This is the smaller value of affiliations the domain of the ligand and the receptor is neutral to the class and this as Complementarity of the domains used. The overall complementarity of the domains again results as an average of the complementarity of the individual combinations of α level amounts.

C. Geometric hashing algorithm

A stable complex structure should be complemented by a variety of Domain combinations whose geometric location is a simultaneous overlap of the Surfaces of these domains allows to mark. The to find this Domain Combinations, further referred to as Geometric Hashing- Algorithm designated by Schwartz et al. [7,8] for the robot control developed. The group of Nussinov and Wolfson [9-12] used this algorithm for a chemical question, wherein based on the location of the C α atoms of the amino acids Structural comparison of proteins with low sequence homology was made. The Extension to the one by Lin et al. [13] proposed subdivision of the molecular Surface led to a successful docking algorithm [14,15]. The geometric hashing algorithm is based on the comparison of the arrangement of indexed points in three-dimensional space and consists of two sub-steps [16,17]: For comparison, only the coordinates of the points as well as the points assigned to the points Index, which describes the properties of the point used. In the first step (Preprocessing) becomes for each molecule an independent of the transformation internal Coordinate system defined. For each reference point then the new coordinates in calculated by this system. In the second step (Recognition) is by comparison of the internal Coordinates the largest number of possible complementary domain combinations certainly.

IV. Discussion

The docking algorithm is developed is based on the three-dimensional structural data of biochemical complex partners the active center identified a receptor and suggestions for possible binding modes of a ligand in this active center. The algorithm is based on the model of a molecular surface. Every point of this Surface is characterized by the topographic properties, the electrostatic potential, the Lipophilicity and the ability to form hydrogen bonds characterized. The Properties are mapped to structural variables that help their similarity between surface points can be calculated. This similarity is used to divide the surface into larger areas with similar properties. The limits of The resulting domains, called domains, will have a membership function which ranges from a value of 1 for full relationship to one Value of 0 for non-domain points is in progress. This makes it possible that Surface points belong to multiple domains and that the domains overlap. It also allows the size of the domains to be matched.

With the help of the Geometric Hashing algorithm [16,17] possible complex structures certainly. For this, the surface domains are characterized by characteristic points and a associated normal vector described and the three-dimensional position of this characteristic points of both molecules compared. The characteristic Points are used as centers of the topographical properties, the electrostatic potential and the lipophilicity-partitioned domains as well as the location defined hydrogen bond partner on the molecular surfaces.

References

[1] Bernstein, F.; Koetzle, T.; Williams, G.; Meyer, E. jr.; Brice, M.; Rodgers, J.;Kennard, J.; Shimanouchi, T.; and Tasumi, M.; The Protein Data Bank: A Computer- Based Archival File for Macromolecular Structures; J.Mol.Biol. (1977); 112; 535-542.

[2] Connolly, M. L.; Solvent Accessible Surfaces of Proteins and Nucleic Acids; Science (1983); 221; 709-713.

[3] Lin, S. L.; Nussinov, R.; Fischer, D.; and Wolfson, H. J.; Molecular Surface Representations by Sparse Critical Points; Proteins (1994); 18; 94-101.

[4] Fischer, D.; Lin, S. L.; Wolfson, H. J.; and Nussinov, R.; A Geometry-Based Suite of Molecular Docking Processes; J.Mol.Biol. (1995); 248; 459-477.

[5] Heiden, W. and Brickmann, J.; Segmentation of Protein Surfaces Using Fuzzy Logic; J.Mol.Graphics (1994); 12; 106-115.

[6] Heiden, W.; Methoden zur computerunterstützten Untersuchung selektiver Oberflächeneigenschaften von Proteinen; Dissertation; Technische Hochschule Darmstadt (1993).

[7] Lamdan, Y.; Schwartz, J. T.; and Wolfson, H. J.; Affine Invariant Model-Based Object Recognition; IEEE Transactions on Robotics and Automation (1990); 6(5); 578-589.

[8] Schwartz, J. T. and Sharir, M.; Identification of Partially Obscured Objects in Two and Three Dimensions by Matching Noisy Characteristic Curves; Int.J.Robotics Res. (1987); 6(2); 29-44.

[9] Fischer, D.; Bachar, O.; Nussinov, R.; and Wolfson, H. J.; An Efficient Automated Computer Vision Based Technique for Detection of Three Dimensional Structural Motifs in Proteins; J.Biomol.Struct.& Dyn. (1992); 9(4); 769-789.

[10] Fischer, D.; Wolfson, H. J.; and Nussinov, R.; Spatial, Sequence-Order-Independent Structural Comparison of Proteins: Evolutionary Implications; J.Biomol.Struct.& Dyn. (1993); 11(2); 367-380.

[11] Fischer, D.; Wolfson, H. J.; Lin, S. L.; and Nussinov, R.; Three-Dimensional, Sequence Order-Independent Structural Comparison of a Serine Protease Against the Crystallographic Database Reveals Active Site Similarities: Potential Implications to Evolution and to Protein Folding; Protein Sci. (1994); 3; 769-778.

[12] Fischer, D.; Tsai, C.-J.; Nussinov, R.; and Wolfson, H. J.; A 3D Sequence-Independent Representation of the Protein Data Bank; Protein Eng. (1995); 8(10); 981-997.

[13] Tsai, C. J.; Lin, S. L.; Wolfson, H. J.; and Nussinov, R.; A Dataset of Protein-Protein Interfaces Generated With a Sequence-Order-Independent Comparison Technique; J.Mol.Biol. (1996); 260; 604-620.

[14] Yeates, T. O.; Algorithms for Evaluating the Long-Range Accessibility of Protein Surfaces; J.Mol.Biol. (1995); 249; 804-815.

[15] Lin, S. L. and Nussinov, R.; Molecular Recognition Via Face Center Representation of Molecular Surface; J.Mol.Graphics (1996); 14; 78-90.

[16] Lamdan, Y.; Schwartz, J. T.; and Wolfson, H. J.; Affine Invariant Model-Based Object Recognition; IEEE Transactions on Robotics and Automation (1990); 6(5); 578-589.

[17] Schwartz, J. T. and Sharir, M.; Identification of Partially Obscured Objects in Two and Three Dimensions by Matching Noisy Characteristic Curves; Int.J.Robotics Res. (1987); 6(2); 29-44.

ISBN: 978-988-14047-9-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)