ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
“Antonio Ruberti”
CONSIGLIO NAZIONALE DELLE RICERCHE

G. Liuzzi, P. Bertolazzi, C. Guerra

A GLOBAL OPTIMIZATION ALGORITHM FOR PROTEIN STRUCTURE ALIGNMENT

R. 08-04, 2008

Paola Bertolazzi – Istituto di Analisi dei Sistemi ed Informatica del CNR, viale Manzoni 30 - 00185 Roma, Italy. Email: `bertola@iasi.cnr.it`.

C. Guerra – Dipartimento di Ingegneria Informatica, Universitá di Padova, Italy and Geaorgia Institute of Technology, Atlanta, USA. Email: `Concettina.Guerra@dei.unipd.it`.

G. Liuzzi – Istituto di Analisi dei Sistemi ed Informatica del CNR, viale Manzoni 30 - 00185 Roma, Italy. Email: `liuzzi@iasi.cnr.it`.

ISSN: 1128–3378

Collana dei Rapporti dell'Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", CNR

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: `iasi@iasi.cnr.it`
URL: `http://www.iasi.cnr.it`

Abstract

In this paper we propose a new method for protein structure alignment based on global optimization techniques. There is evidence [23, 9] that proteins with similar shape typically perform similar functions. Given the three-dimensional structures of two proteins, we are interested in finding the isometric transformation (rotation plus translation) that best aligns the two structures. We draw our inspiration from the well-known Iterative Closest Point (ICP) method for three-dimensional (3D) shapes registration. Our algorithm globally minimizes a dissimilarity measure between the two structures. The reported computational experience and comparison show viability of the proposed approach.

Abstract

In this paper we propose a new method for protein structure alignment based on global optimization techniques. There is evidence [23, 9] that proteins with similar shape typically perform similar functions. Given the three-dimensional structures of two proteins, we are interested in finding the isometric transformation (rotation plus translation) that best aligns the two structures. We draw our inspiration from the well-known Iterative Closest Point (ICP) method for three-dimensional (3D) shapes registration. Our algorithm globally minimizes a dissimilarity measure between the two structures. The reported computational experience and comparison show viability of the proposed approach.

1. Introduction

The Protein structure alignment problem consists in finding residue pairs occupying spatially equivalent positions. The problem has been extensively studied [10, 17, 16, 24]. Almost all the algorithms for protein structure alignment search for an alignment between C_α atoms of the proteins. A variety of protein structure alignment algorithms have been proposed and used in the literature [22, 25, 15, 26, 10, 17, 1, 12].

Most of these methods require the solution of a 3D matching or registration problem which is a well-known problem in computer graphics and computer vision. It can be formulated as follows: given two sets A and B of points, find the subset of B with highest "*similarity*" to set A.

There are various ways of defining the similarity between two point sets in 3D space and these lead to the proposal of different algorithms (closest point distance [2], Hausdorff distance [8], bottleneck distance [1, 6], root mean square distance [1]).

Furthermore, in [12] a deep analysis and comparison among many of the principal structural alignment algorithms is presented

Our alignment problem is similar to the aforementioned 3D matching problem. Indeed, given two sets A and B of points, corresponding to the C_α atoms of two proteins backbone, we search for the subset of B which is most *similar* to set A. One possible way to solve the problem is by using the well-known Iterative Closest Point (ICP) algorithm [2, 27] from which we draw our inspiration. Indeed, we adopt a similar approach, namely, we search for the isometric transformation which best superimposes the two given protein structures making use of a new dissimilarity measure along with a global optimization approach.

In order to solve the problem, we first introduce a new definition of dissimilarity. The dissimilarity measure that we propose is based on the solution of an "*Asymmetric Assignment Problem*" on a particular bipartite graph. Then, we propose the use of a global optimization algorithm to find the isometric transformation which yields the best similarity between the two point sets.

The global optimization algorithm that we use [18, 3, 4] belongs to the class of controlled random search methods. This kind of methods, though heuristic in nature, are very efficient and reliable for the global minimization of a nonlinear multivariate function of several variables. In the past years controlled random search algorithms have been successfully used to solve many real world problems, see for instance [5, 13, 14, 21, 20, 19].

The paper is organized as follows. In Section 2 we introduce some useful notations and assumptions to be used in the paper. Section 3 is devoted to the introduction of a class of iterative methods for structure alignment, namely the Iterative Point-to-Point Algorithm class. In Section 4 we describe in some details the global optimization algorithm that is used to solve the problem by briefly recalling its structure. In Section 5 we define the new dissimilarity measure based on the solution of an Asymmetric Asignment Problem. Finally, in Section 6 we report some preliminar numerical results and the comparison between our code and the method proposed in [27].

2. Notations and Assumptions

In this section we introduce some notations and assumptions that will be used throughout the paper. Given two protein structures $\mathcal{P}$ and $\mathcal{Q}$, let us denote by P and Q the two finite sets of points corresponding to C_α atoms of the two structures $\mathcal{P}$ and $\mathcal{Q}$, respectively.

Set P is conventionally representative of a query shape while Q defines a reference model shape.

An isometric transformation in three-dimensional space can be defined by means of a unit quaternion $a_r = (a_0, a_1, a_2, a_3)^\top \in \Re^4$ ($\|a_r\| = 1$) and of a translation vector $a_t \in \Re^3$. Let $a^\top = (a_r^\top \; a_t^\top)$ be the transformation defining vector and denote by T_a the corresponding transformation, so that

$$y = T_a(x) = R(a_r)x + a_t$$

for every $x \in \Re^3$, where $R(a_r)$ is the rotation matrix defined by means of the unit quaternion a_r as follows[1]

$$R(a_r) = \begin{pmatrix} a_0^2 + a_1^2 - a_2^2 - a_3^2 & 2(a_1a_2 - a_0a_3) & 2(a_1a_3 + a_0a_2) \\ 2(a_1a_2 + a_0a_3) & a_0^2 + a_2^2 - a_1^2 - a_3^2 & 2(a_2a_3 - a_0a_1) \\ 2(a_1a_3 - a_0a_2) & 2(a_2a_3 + a_0a_1) & a_0^2 + a_3^2 - a_1^2 - a_2^2 \end{pmatrix}$$

Let $\Theta \subset \Re^7$ be the set of all vectors $a \in \Re^7$ defining an isometric transformation in $\Re^3$.

Given a transformation vector $a \in \Theta$, let $T_a(P) = P_a$ denote the set of points obtained applying transformation T_a to every point of P, that is

$$T_a(P) = P_a = \{y : y = R(a_r)p + a_t, \; \forall \, p \in P\}.$$

Let $\psi : P \to Q$ denote a point to point function that associates to every point of p a point of Q. Since, as assumed above, P and Q are finite sets, the class Ψ of all functions ψ has finite cardinality and it results

$$|\Psi| = m^n.$$

A function $\varphi \in \Psi$ is an *assignment* from P to Q if, by definition, it is injective, that is when for every $p_1, p_2 \in P$

$$p_1 \neq p_2 \quad \Rightarrow \quad \varphi(p_1) \neq \varphi(p_2).$$

Let us denote by $\Phi \subseteq \Psi$ the class of all possible assignments from P to Q. Obviously, since P and Q are finite sets, Φ is finite as well and it results

$$|\Phi| = m(m-1)\ldots(m-n+1).$$

Let $\psi \in \Psi$ be a given mapping and a be a vector defining an isometric transformation, then the mean square error function between P and Q is the following

$$f(\psi, a) = \frac{1}{n} \sum_{p \in P} \|\psi(p) - R(a_r)p - a_t\|^2. \tag{1}$$

The structural alignment problem consists in finding a mapping $\psi^* \in \Psi$ of points in P to points in Q and an isometric transformation a^* such that

$$f(\psi^*, a^*) \leq f(\psi, a),$$

for all $\psi \in \Psi$ and $a \in \Theta$.

3. A class of iterative algorithms

In this section we introduce a class of iterative algorithms which generalizes many algorithms for shape registration. One of the best known algorithms for shape registration, namely the Iterative Closest Point Algorithm [2], belongs to this general class of iterative algorithms.

---

[1] Otherwise, a rotation matrix can be expressed in terms of Euler angles as follows

$$R(\alpha, \beta, \gamma) = \begin{pmatrix} c_\beta c_\gamma & s_\alpha s_\beta c_\gamma - c_\alpha s_\gamma & c_\alpha s_\beta c_\gamma + s_\alpha s_\gamma \\ c_\beta s_\gamma & s_\alpha s_\beta s_\gamma + c_\alpha c_\gamma & c_\alpha s_\beta s_\gamma - s_\alpha c_\gamma \\ -s_\beta & s_\alpha c_\beta & c_\alpha c_\beta \end{pmatrix}$$

where $s_\alpha = \sin\alpha$, $c_\alpha = \cos\alpha$ and accordingly.

All of these algorithms stem from the idea that, once a mapping $\bar{\psi} \in \Psi$ is fixed, it is possible to compute the exact minimizer of the function $f(\bar{\psi}, a)$ with respect to the vector $a \in \Theta$ defining the isometric transformation. Let $a(\bar{\psi})^*$ be the minimizer of $f(\bar{\psi}, a)$, that is

$$a(\bar{\psi})^* = \arg\min_{a \in \Theta} f(\bar{\psi}, a).$$

A closed-form expression for $a(\bar{\psi})^*$ has been given in [11] where we refer the interested reader for the relevant details.

Hence, the problem implicitly considered by an iterative algorithm is the following two-level optimization problem

$$\begin{array}{l} \min_{\psi} f(\psi, a) \\ s.t.\ \psi \in \Psi \\ \qquad a = \arg\min_{a \in \Theta} f(\psi, a). \end{array} \qquad (2)$$

In the following we report the general scheme of an Iterative Point-to-Point (IPP) Algorithm

IPP Algorithm.

0) Set $P_0 = P$, $a_0 = (1, 0, 0, 0, 0, 0, 0)^\top$, $k = 0$.

1) Choose a mapping $\psi_k \in \Psi$ according to some criterion and Compute $f_k = f(\psi_k, a_k)$.

2) Compute the registration $a_{k+1} = a(\psi_k)^*$.

3) Apply the registration $P_{k+1} = T_{a_{k+1}}(P_0)$.

4) Set $k = k + 1$ and go to Step 1).

The general IPP Algorithm manages to tackle the twofold problem of computing a mapping $\psi : P \to Q$ and, once given ψ, computing the optimal transformation $a(\psi)^*$, by iteratively performing steps 1) and 2). In particular, Step 1) is devoted to the choice of a mapping ψ of points belonging to P to points of Q. Step 2), on the other hand, is devoted to the computation of $a(\psi)^*$ that is the new estimate of the optimal transformation once fixed the mapping ψ.

If, in particular, at every iteration the mapping ψ_k at Step 1) is chosen to be

$$\psi_k = \arg\min_{\psi \in \Psi} \frac{1}{n} \sum_{p \in P} \|\psi(p) - T_{a_k}(p)\|^2,$$

we obtain the well-known *Iterative Closest Point* (ICP) Algorithm of [2]. It is worth noting that, in this particular case, the mapping ψ used at Step 1) is the one that associates every point of set $T_{a_k}(P)$ to the closest point in Q, from which the name *Closest Point* mapping and hence Iterative Closest Point Algorithm.

In [2] the convergence of Algorithm ICP has been proved. In particular, the sequence $\{f_k\}$ of mean square error function values, with

$$f_k = \frac{1}{n} \sum_{p \in P} \|\psi_k(p) - T_{a_k}(p)\|^2,$$

is proved to be monotonically decreasing thus, since $f_k \geq 0$, convergent to a lower bound $\bar{f}$.

As it is stated in [2], where ICP has been originally proposed, the method converges to a solution which is a local minimum of the considered distance function, that is to say that ICP Algorithm solves the two-level problem (2) to local optimality. Indeed, it has been shown [2] that the final transformation $\bar{a}$ and mapping $\bar{\psi}$ spotted by the Algorithm ICP heavily depend on the initial relative positioning of sets P and Q.

4. A global optimization algorithm

In this section we discuss the use of a global optimization algorithm for the solution of the shape registration problem. To this aim, we note that it is possible to formulate the problem in a complementary way with respect to Problem (2). More in particular, let us think at the inner-level problem as the one defining the mapping function ψ (instead of the transformation a) once fixed the transformation vector $a \in \Theta$. Namely, we can define the following two-level optimization problem

$$\begin{array}{l} \min_a f(\psi, a) \\ s.t.\ a \in \Theta \\ \qquad \psi = \arg\min_{\psi \in \Psi} f(\psi, a). \end{array} \tag{3}$$

The above problem (3) can be reduced to a one-level optimization problem by considering that for every vector $a \in \Theta$, the lower-level problem of (3) admits a global solution, which we denote by

$$\psi(a) = \arg\min_{\psi \in \Psi} f(\psi, a) \tag{4}$$

which is the closest point mapping that we introduced in the latter section. Hence, Problem (3) can be equivalently stated as

$$\min_{a \in \Theta} g(a) \tag{5}$$

where $g(a) = f(\psi(a), a)$. Every global solution a^* of (5) is, by definition, a solution such that $f(\psi(a^*), a^*) \leq f(\psi(a), a)$, for all $a \in \Theta$.

Observe that the computation of function g requires the computation of the optimal mapping $\psi(a)$, that is, the solution of Problem 4. This latter problem can be solved with a time complexity $O(nm)$ in the worst case [2] which can be relevant for n and m large. Moreover, due to its definition, $g(a)$ is a non-smooth (Lipschitz) continuous function and its derivatives are not available. Indeed, for the minimization of function $g(a)$ we cannot neither directly use its derivatives nor approximate them through finite differences since this would require too much time and produce numerical derivatives which are unreliable because of the non-smoothness of function g.

On the basis of the above observation we propose the use of a controlled random search method for the solution of Problem (5).

In the following we briefly recall the global optimization algorithm that we use and which was originally proposed in [18] and successively improved in [4]. It is a population based algorithm in the sense that throughout the entire optimization process a population of points (i.e. vectors $w \in \Re^N$, $N = 6$) is maintained and iteratively updated in such a way that they cluster around the global minima of the objective function. Roughly speaking, the method is composed of two distinct and consecutive phases: a global phase and a local phase. During the so-called global phase an initial population of points is generated by randomly sampling a sufficiently large set of points over some feasible domain. Then, at every iteration of the local phase, a new point is generated and the population is updated if this new point improves on the worst point of the population. More in details, the algorithm can be described by the following steps.

1. `initialization` The objective function is sampled on a set S of M points randomly chosen within some feasible domain $\mathcal{F}$ strictly containing the global minimizer.

2. `stopping criterion` If the maximum and minimum values of the objective function over S are sufficiently close to each other, namely

   $$f_{\max} - f_{\min} < \epsilon,$$

   where

   $$f_{\max} = \max_{w \in S} f(w), \quad f_{\min} = \min_{w \in S} f(w)$$

   then STOP.

3. `search phase` $N+1$ points are randomly chosen on set S, where N is the dimension of the problem ($N=6$ in our case). Then,
   (a) the *weighted* centroid w_c of the $N+1$ points is computed;
   (b) the new trial point $\tilde{w}$ is computed by doing a *weighted reflection* of the centroid onto the worst point among the selected $N+1$ points. Namely, let $w^\dagger$ be the worst point, then

   $$\tilde{w} = (1+\alpha)w_c - \alpha w^\dagger,$$

   where $\alpha \in (0,1)$ is a reflection parameter and is iteratively updated during the optimization process [4] in such a way that its value tends to zero as the iteration count increases and the difference $f_{\max} - f_{\min}$ decreases.
4. `updating phase` If the objective function value on the new point $\tilde{w}$ improves on the maximum function value over S, then the set S is updated by adding the new point and discarding the worst one. Otherwise set S is left unchanged and the new point is discarded. The algorithm continues iterating through steps 2–4.

5. A new dissimilarity measure

In this section we propose a new dissimilarity measure between two given sets of points of the two proteins. This measure is based on a distance other than the closest point distance mentioned in the preceding section and that is used in the definition of Algorithm ICP.

In particular, it can be noted that, using the closest point distance, it can happen that different point of set $T_a(P)$ are mapped to the same point of set Q. This, in turn can yield a distance value which is small just because many points are all mapped to the same closest point.

In order to avoid this undesirable effect, let us consider function $f(\varphi, a)$ defined in (1) where $\varphi \in \Phi$ is an assignment function and let, for every a, $\varphi(a)$ be a global solution of problem

$$\min_{\varphi \in \Phi} f(\varphi, a). \tag{6}$$

Problem (6) can be formulated as a 0,1-optimization problem and is, indeed, a well-known combinatorial optimization problem that goes under the name of *Asymmetric Assignment Problem* (AAP).

In particular, let $G(P,Q,E)$, $E = P \times Q$, be the bipartite directed graph characterized by the two set of nodes P and Q and with edge set given by the cartesian product of P times Q. For every pair $e=(p,q) \in E$, define

$$c_e = \|q - T_a(p)\|^2.$$

Let $s \in \{0,1\}^{|E|}$ be the edge incidence vector and consider the following minimum cost assignment problem

$$\begin{aligned} &\min_s c^\top s \\ &\sum_{e \in \delta^+(p)} s_e = 1, \quad \forall p \in P \\ &\sum_{e \in \delta^-(q)} s_e \le 1, \quad \forall q \in Q \\ &s \in \{0,1\}^{|E|}, \end{aligned} \tag{7}$$

where, $\delta^+(p)$ and $\delta^-(q)$, are the sets of edges leaving node p and, respectively, entering in node q.

Note that the constraints of Problem (7) require each node $p \in P$ to be assigned to exactly one node $q \in Q$ and each node $q \in Q$ to be assigned to at most one node $p \in P$, which is why Problem (7) is known as *Asymmetric Assignment Problem.*

Naturally, it results

$$f(\varphi(a), a) = c^\top s^*$$

where s^* is the optimal solution of Problem (7).

Problem (7), and hence (6), can be solved very efficiently by *ad-hoc* codes that have time complexity $O(\sqrt{n}m\log(n\mathcal{C}(T_a)))$ where $\mathcal{C}(T_a) = \max_{p\in P, q\in Q}\{\|q - T_a(p)\|\}$, see for instance [7].

We are now able to define our new dissimilarity measure, that we call *matching distance.*

Definition 5.1. *Given an isometric transformation $a \in \Theta$ and two distinct sets of points P and Q, the* matching distance *between $T_a(P)$ and Q is given by $f(\varphi(a), a)$.*

Reasoning as in the preceding section, we can now search for a global solution of problem (5) where now $g(a) = f(\varphi(a), a)$.

6. Preliminary Results

In this section we report a preliminar experimentation of our global optimization code on a set of protein pairs. The set of protein pairs is the same as that used in [27] where a modification of the well-known ICP Algorithm is proposed for the structural alignment problem.

The table that follows report the results obtained by our code compared with the results of paper [27] and those obtained using the original ICP Algorithm [2].

Protein pairs		GlobProt	CMIs + ICP	ICP
1bge-B	2gmf-A	101/2.1	83/2.1	19/2.1
1cew-I	1mol-A	40/1.6 (68/2.2)	68/1.6	13/1.6
1cid	2rhe	72/1.6 (80/1.7)	80/1.6	0/1.6 (80/10.0)
1crl	1ede	251/2.3	193/2.3	34/2.3
1fxi-A	1ubq	58/1.6	54/1.6	8/1.6
1ten	3hhr-B	38/1.5 (76/2.2)	76/1.5	0/1.5 (76/18.2)
1tie	4fgf	109/2.2	80/2.2	45/2.2
2sim	1nsb-A	283/2.3	201/2.3	149/2.3
2aza-A	1paz	85/2.1	77/2.1	38/2.1
3hla-B	2rhe	83/2.1	59/2.1	20/2.1

For each pair of proteins we report the number of matched C_α atoms along with the computed mean square error. For a better comparison with the results from [27] of Algorithm CMIs+ICP and with the original ICP Algorithm, for some protein pairs, we also report the value of the mean square error that would be obtained if we required that the number of matched C_α atoms is exactly that obtained by CMIs+ICP.

A. Closed-form solution of the absolute orientation problem

In [11] has been given a closed-form expression for the minimizer of $f(a, t)$ which we now briefly recall.

Let
$$\mu_P = \frac{1}{n}\sum_{p\in P} p, \quad \text{and} \quad \mu_Y = \frac{1}{n}\sum_{p\in P} \psi(p)$$

be the centers of mass of point sets P and Y. The cross-covariance of the sets P and Y is given by
$$\Sigma = \frac{1}{n}\sum_{p\in P} p\psi(p)^\top - \mu_P\mu_Y^\top.$$

Consider the 4×4 matrix
$$Q(\Sigma) = \begin{pmatrix} \mathrm{tr}(\Sigma) & \Delta^\top \\ \Delta & \Sigma + \Sigma^\top - \mathrm{tr}(\Sigma)I \end{pmatrix},$$

where I is the 3×3 identity matrix and $\Delta = (B_{23}, B_{31}, B_{12})^\top$ with $B = \Sigma - \Sigma^\top$.

It results ([11, 2]) that a^* is the unit eigenvector corresponding to the maximum eigenvalue of $Q(\Sigma)$ and
$$t^* = \mu_Y - R(a^*)\mu_P.$$

References

[1] T. Akutsu "*Protein structure alignment using dynamic programming and iterative improvement*", IEICE Trans. Inf. and Syst., Vol. E78-D, pp. 1–8 (1996)

[2] P.J. Besl, and N.D. McKay "*A method for registration of 3-D shapes*", IEEE Trans. on Pattern Analysis and Mach. Intelligence, Vol. 14, pp. 239–255 (1992)

[3] P. Brachetti, M. De Felice Ciccoli, G. Di Pillo, and S. Lucidi "*A new version of the Price's algorithm for global optimization*", Journal of Global Optimization, Vol. 10, pp. 165–184 (1997)

[4] L. Cirio, S. Lucidi, F. Parasiliti, and M. Villani "*A global optimization approach for the synchronous motors design by finite element analysis*", International Journal of Applied Electromagnetics and Mechanics, Vol. 16, pp. 13–27 (2002)

[5] A. Daidone, F. Parasiliti, M. Villani, and S. Lucidi "*A new method for the design optimization of three-phase induction motors*", IEEE Trans. on Magnetics, Vol. 34, pp. 2932–2935 (1998)

[6] A. Efrat, A. Itai, and M.J. Katz "*Geometry helps in bottleneck matching and related problems*", Algorithmica, Vol. 31, pp. 1–28 (2001)

[7] A.V. Goldberg and R. Kennedy "*An Efficient Cost Scaling Algorithm for the Assignment Problem*", Mathematical Programming, Vol. 71, pp. 153-178 (1995)

[8] M.T. Goodrich, J.S.B. Mitchell, and M.W. Orletsky "*Practical methods for approximate geometric pettern matching under rigid motions*", Proc. of the 10th Ann. Symp. on Computational Geometry, ACM Press, pp. 103–112 (1994)

[9] M.E.Bock, C. Garutti, and C. Guerra "*Discovery of Similar Regions on Protein Surfaces*", Journal of Computational Biology, Vol. 14, pp. 285-299 (2007)

[10] L. Holm, C. Onzounis, C. Sander, G. Tuparev, and G. Vriend "*A database of protein structure families with common folding motifs*", Protein Science, Vol. 1, pp. 1691–1698 (1992)

[11] B.K.P. Horn "*Closed-form solution of absolute orientation using unit quaternions*", Journal of the Optical Society of America A, Vol. 4, pp. 629-642 (1987)

[12] R. Kolodny, P. Koehl and M. Levitt "*Comprehensive Evaluation of Protein Structure Alignment Methods: Scoring by Geometric Measures*" J. Mol. Biol., 346, pp.1173–1188 (2005)

[13] G. Liuzzi, S. Lucidi, V. Piccialli and A. Sotgiu "*A magnetic resonance device designed via global optimization techniques*", Math. Programming Ser. B, Vol. 101, pp. 339–364 (2004)

[14] G. Liuzzi, S. Lucidi, F. Parasiliti, and M. Villani "*Multi-objective Optimization Techniques for the Design of Induction Motors*", IEEE Trans. on Magnetics, Vol. 39, pp.1261-1264 (2003)

[15] R. Nussinov, and H.J. Wolfson "*Efficient detection of three-dimensional structural motifs in biological macromolecules by computer vision techniques*", Proc. National Academy of Science (USA), Vol. 88, pp. 10495–10499 (1991)

[16] C.A. Orengo, D.T. Jones, and J.M. Thornton "*Protein superfamilies and domain superfolds*", Nature, Vol. 372, pp. 631–634 (1994)

[17] S. Pascarella, and P. Argos "*A data bank merging related protein structures and sequences*", Protein Engineering, Vol. 5, pp. 121–137 (1992)

[18] W.L. Price "*A controlled random search procedure for global optimization*", in Towards Global Optimization 2, L. Dixon and G. Szego, eds., North-Holland, Amsterdam (1978)

[19] W.L. Price "*Global optimization by controlled random search*", Journal of Optimization Theory and Applications, Vol. 40, pp. 333–348 (1983)

[20] W.L. Price "*Global optimization algorithms for a CAD workstation*", Journal of Optimization Theory and Applications, Vol. 55, pp. 133–146 (1983)

[21] W.L. Price and F. Woodhams "*Optimising accelerator for CAD workstations*", IEEE proceedings, Vol. 135, pp. 214–221 (1988)

[22] S.T. Rao, and M.G. Rossmann "*Comparison of supersecondary structures in proteins*", J. Molecular Biology, Vol. 76, pp. 241–256 (1973)

[23] M. Rosen, S.L. Lin, H.J. Wolfson, and R. Nussinov "*Molecular shape comparisons in searches for active sites and functional similarity*", Protein Engineering, Vol. 11, pp. 263-277 (1998)

[24] A. Šali, and J.P. Overington "*Derivation of rules for comparative protein modeling from a database of protein structure alignments*", Protein Science, Vol. 3, pp. 1582–1596 (1994)

[25] W.R. Taylor, and C.A. Orengo "*Protein Sstructure alignment*", J. Molecular Biology, Vol. 208, pp. 1–22 (1989)

[26] G. Vriend, and C. Sander "*Detection of common three-dimensional substructures in proteins*", PROTEINS: Structure, Function and Genetics, Vol. 11, pp. 52–58 (1991)

[27] D. Xu, H. Li and T. Gu "*Protein Structure Superposition by Curve Moment Invariants and Iterative Closest Point*" The 1st International Conference on Bioinformatics and Biomedical Engineering 6, 25–28 (2007).

—————— Unused references ——————————

[28] M. Comin, C. Guerra, and G. Zanotti "*PROuST: A comparison method of three-dimensional structures of proteins using indexing techniques*", Journal of Computational Biology, Vol. 11, pp. 1061-1072 (2004)

[29] D. Fischer, H.J. Wolfson, and R. Nussinov "*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 Science, Vol. 3, pp. 769–778 (1994)

[30] A. G. Murzin, S. E. Brenner, T. Hubbard, C. Chothia "*SCOP: a structural classification of proteins database for the investigation of sequences and structures.* " J. Mol. Biol. 247, 536–540 (1995).

[31] S. Schmitt, D. Kuhn, and G. Klebe "*A new method to detect related function among proteins independent of sequence and fold homology*", Journal of Molecular Biology, Vol. 323, pp. 387–406 (2002)

[32] A. Alesker, R. Nussinov, H.J. Wolfson "*Detection of non-topological motifs in protein structures*“ Protein Engineering, Vol. 9, 5, pp. 1103–1119 (1996)

[33] Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne "*The Protein Data Bank*" Nucleic Acids Research, Vol. 28, pp. 235–242 (2000)

[34] C. Ferrari, C. Guerra "*Geometric methods for protein structure comparison*" Protein Structure Analysis and Design, (C. Guerra, S. Istrail, eds.), Lecture Notes in Bioinformatics, Springer-Verlag. (2003)

[35] H.M. Grindley, P.J. Artymiuk, D.W. Rice and P. Willett "*Identification of tertiary structure resemblance in proteins using a maxima common subgraph isomorphism algorithm*" J. Mol. Biol., Vol. 229, pp. 707–721 (1993)

[36] L. Holm, C. Sander "*Protein structure comparison by alignment of distance matrices*" J. Mol. Biol., 233, pp.123–138 (1993)

[37] L. Holm, C. Sander "*DaliLite workbench for protein structure comparison*" Bionformatics, Vol. 16, pp. 566–567 (2000)

[38] G. Lancia, R. Carr, B. Walenz, S. Istrail "*Optimal PDB structure alignments: A Branch-and-Cut algorithm for the maximum contact map overlap problem.*" *Proc. 5th ACM Research in Computational Biology*, pp. 193–202 (2001)

[39] C. Lemmen, T. Lengauer "*Computational methods for the structural alignment of molecules.*" *J. of Computer-Aided Molecular Design*, Vol. 14, pp. 215–232 (2000)

[40] I.N. Shindyalov, P.E. Bourne "*Protein structure alignment by incremental combinatorial extension (CE) of the optimal path.*" *Protein Eng.*, **11**, Vol. 9, 739–747 (1998).

[41] R. Kolodny, N. Linial "*Approximate protein structural alignment in polynomial time.*" Proc. Natl Acad. Sci. USA, 101, 12201–12206 (2004).

[42] M. Gerstein, M. Levitt "*Comprehensive assessment of automatic structural alignment against a manual standard, the Scop classification of proteins.*" Protein Sci. 7, 445–456 (1998).

[43] J. Xu, F. Jiao and B. Berger "*A Parameterized Algorithm for Protein Structure Alignment*" RECOMB 2006, pp. 488–499.

[44] Y. Zhang, J. Skolnick "*TM-align: a protein structure alignment algorithm based on the TM-score.*" Nucleic Acids Research, vol 33, 2302 (2005).