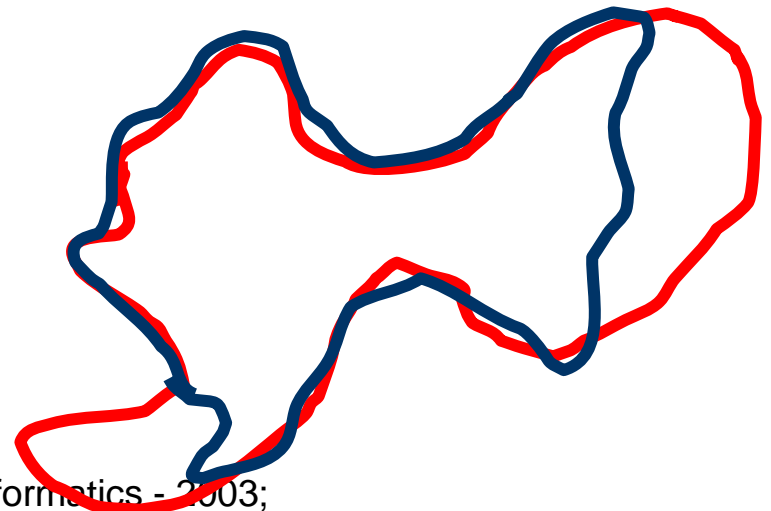
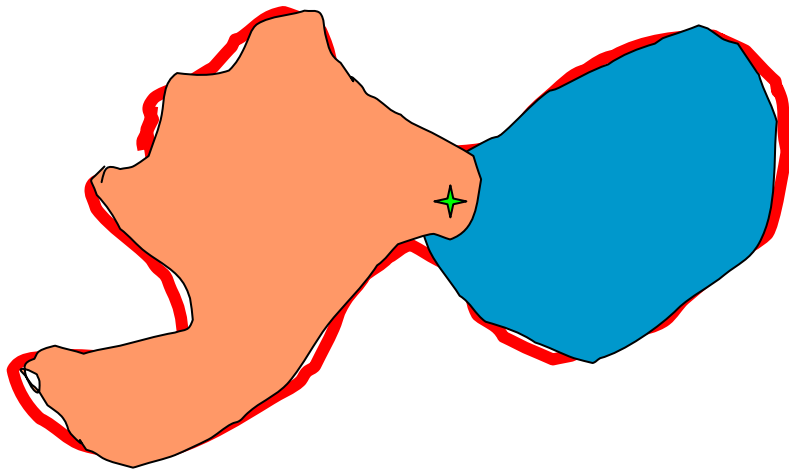
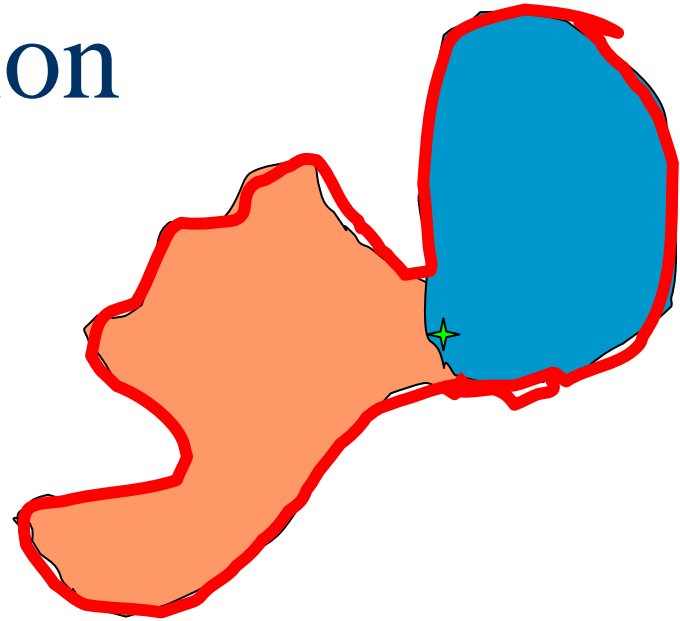
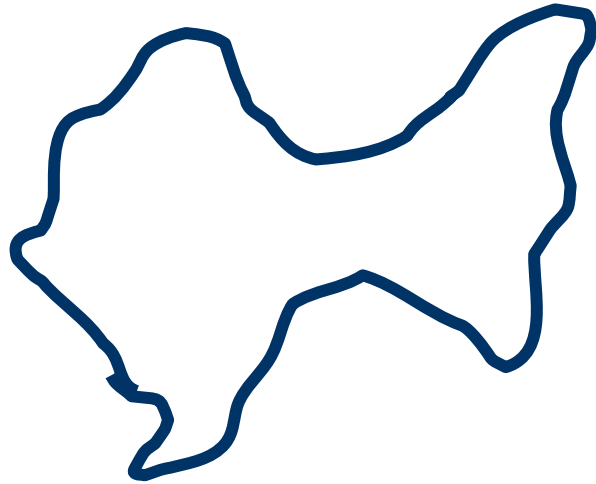


Alignment of Flexible Molecular Structures

Motivation

- Proteins are flexible. One would like to align proteins modulo the flexibility.
- *Hinge* and *shear* protein domain motions (Gerstein, Lesk , Chotia).
- Conformational flexibility in drugs.

Problem definition



Previous Work (1)

- Most work done mainly for drug molecule alignment :
 - Leach and Kuntz (extension of Dock);
 - directed tweak method (Hurst);
 - Lengauer, Rarey, Lemmen (FlexX, FlexS);
 - genetic algorithms (Jones et al.).

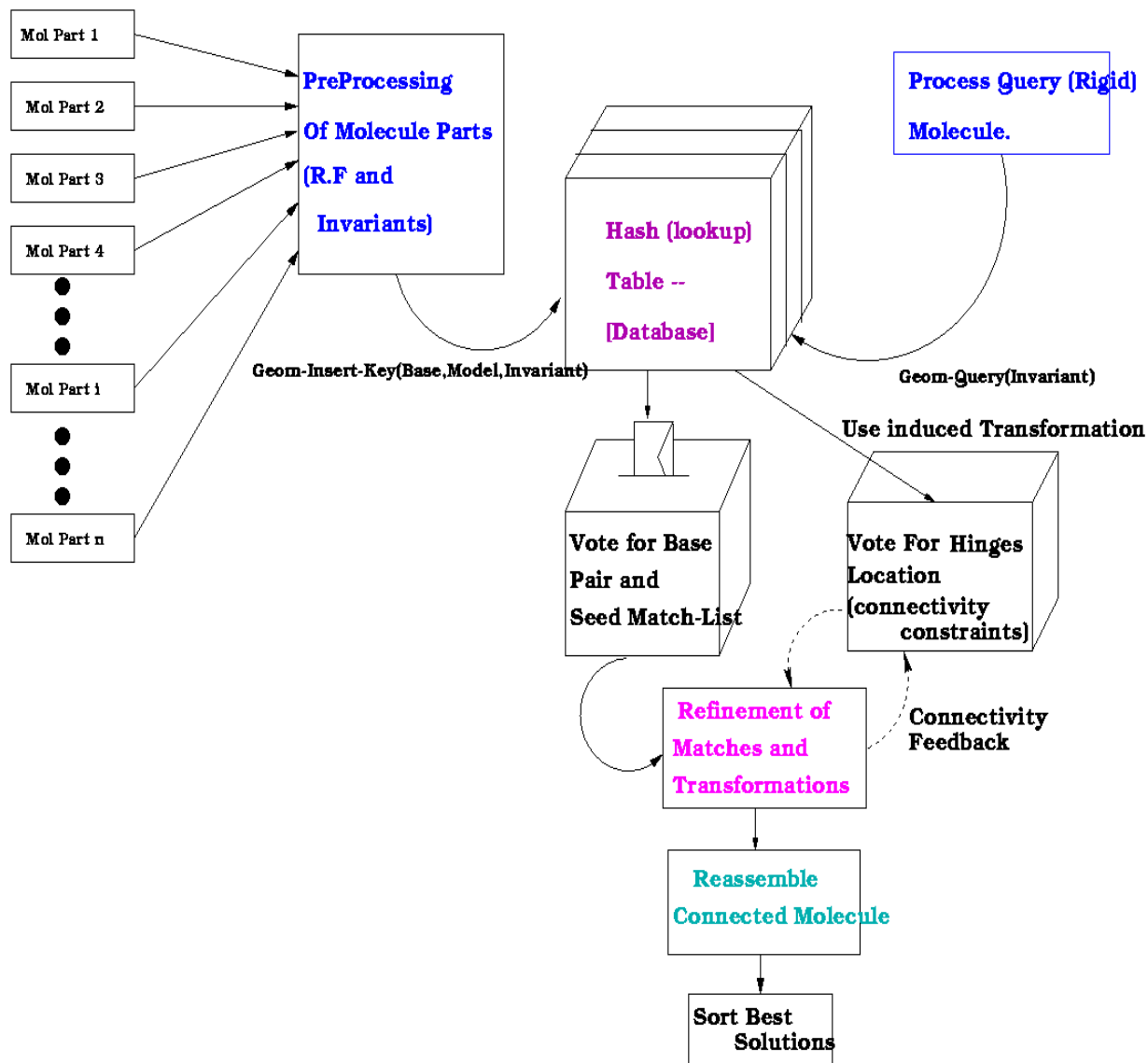
Previous Work (2)

- Sequence **order independent**, flexible protein alignment (*known hinges*) - *Verbitsky et al.*, 1999 - based on “articulated object recognition” in Computer Vision (*Wolfson*, 1991) - Geometric Hashing/Generalized Hough Transform.
- Califano and Rigoutsos - small molecule database search - Geometric Hashing.

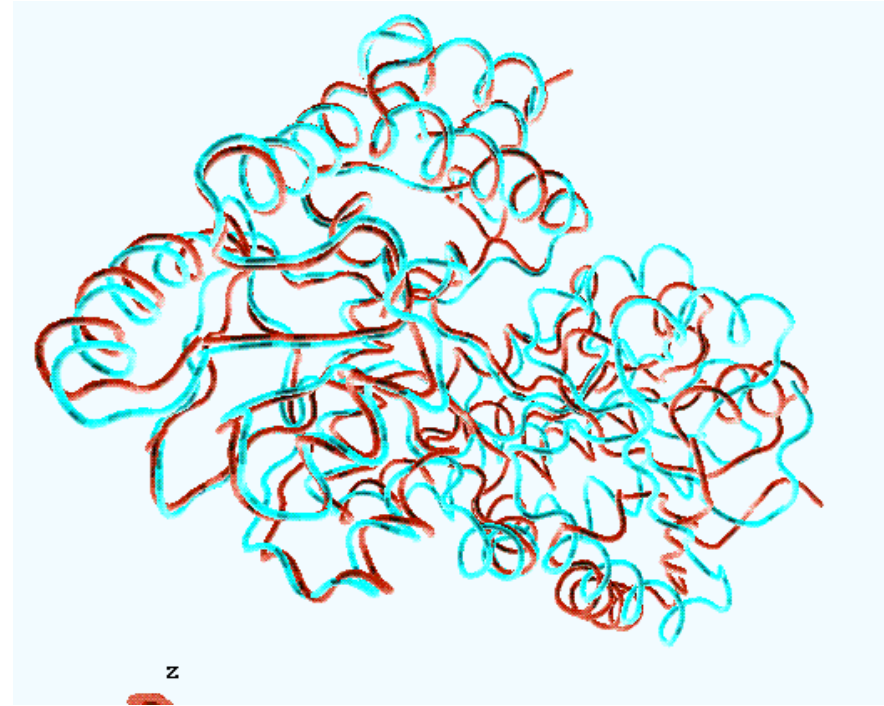
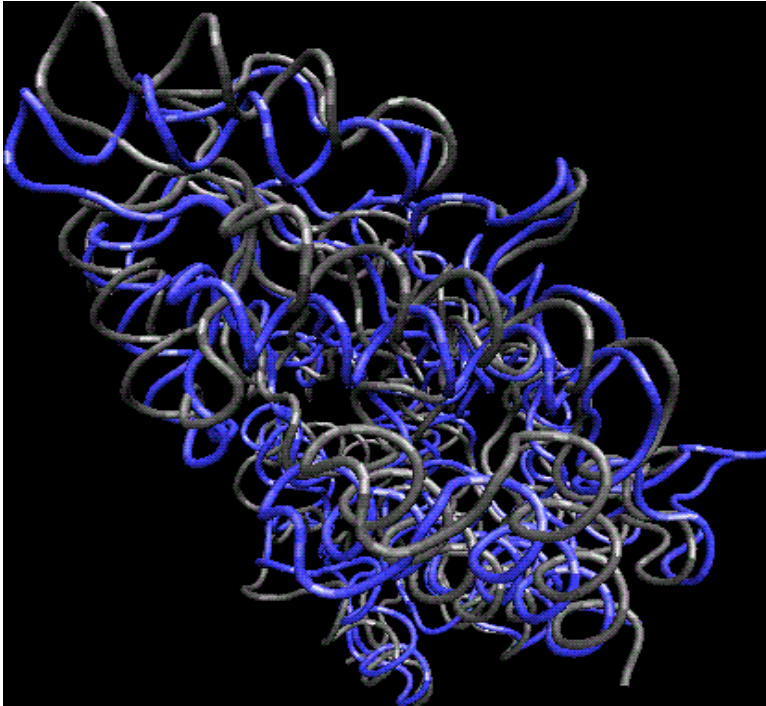
Flexible Geometric Hashing

- Exploit the fact that neighboring parts share the joint - accumulate mutual information at the joint.
- Achieve complexity of the same order of magnitude as in rigid alignment.

Flexible Geometric Hashing



Flexible alignment with several hinges (3)



Alignment of Flexible Protein Structures

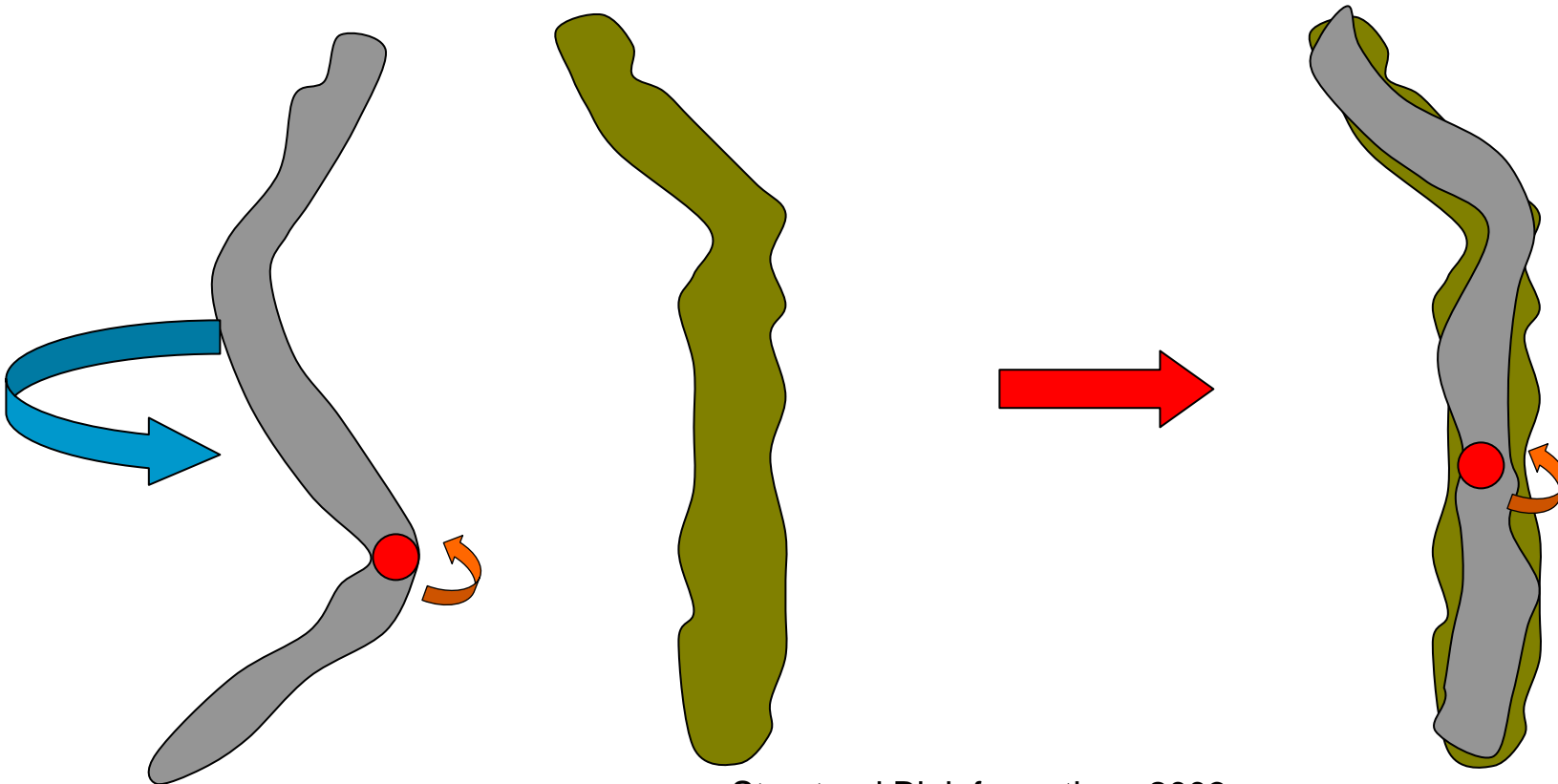
FlexProt **algorithm**

**Maxim Shatsky, Ruth Nussinov,
Haim J. Wolfson**



TEL AVIV University

Motivation

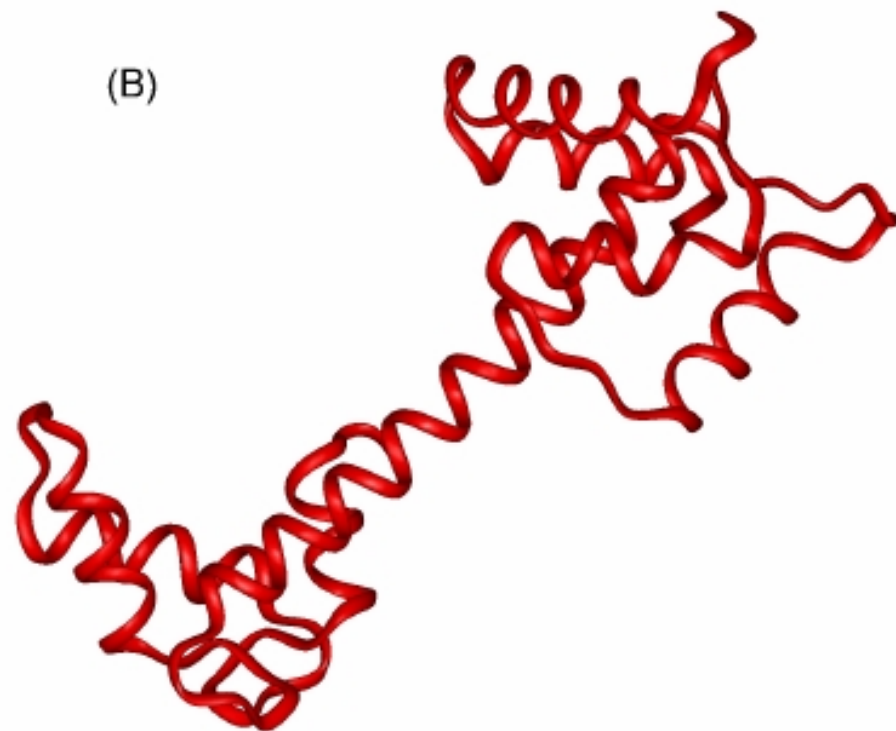


Experimental Results

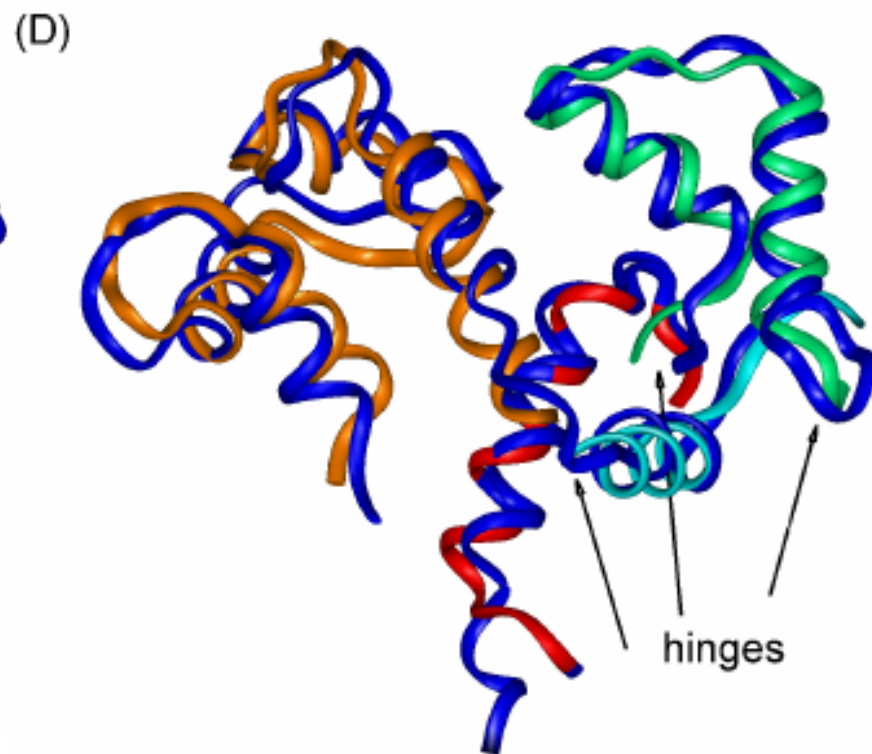
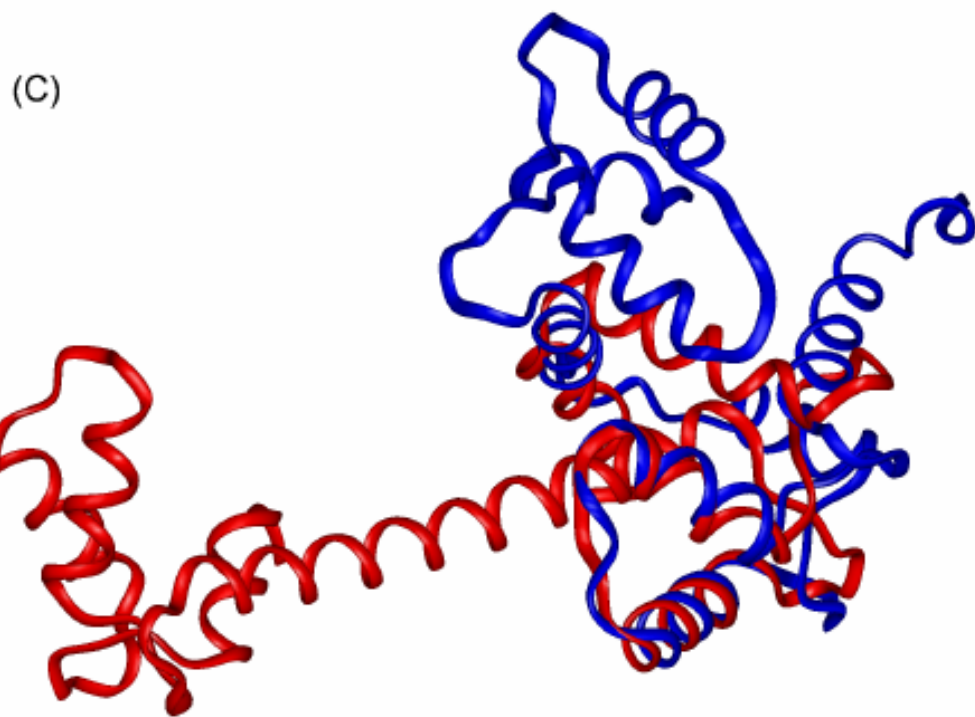
(A)



(B)



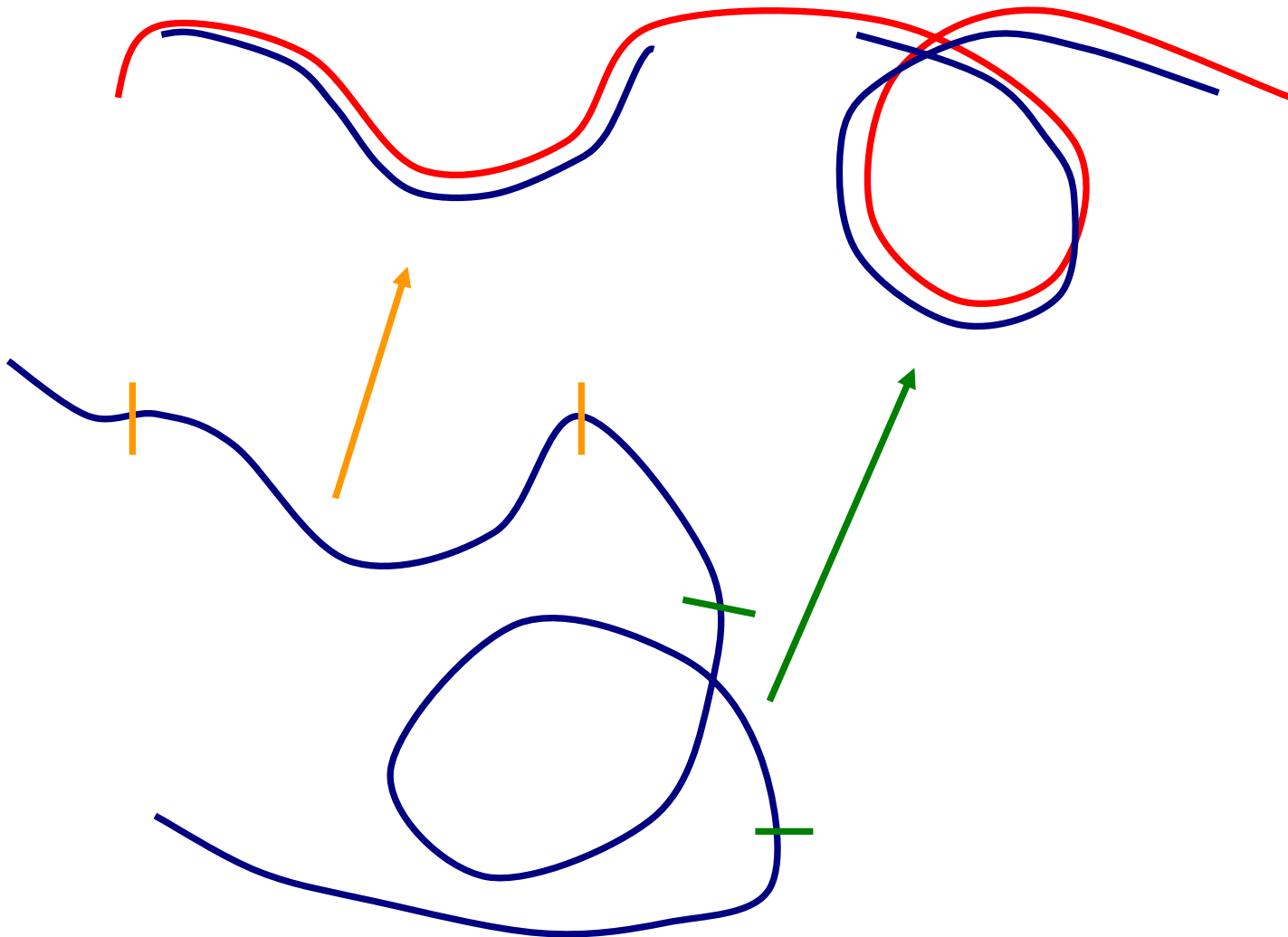
Experimental Results



FlexProt Algorithm

- Input: two protein molecules A and B, each being represented by the sequence of the 3-D coordinates of its C_{α} atoms.
- Task: largest flexible alignment by decomposing the two molecules into a minimal number of rigid fragment pairs having similar 3-D structure.

The Goal



Previous Work - rigid alignment

Sequence order dependent :

- Holm and Sander, 1993 - FSSP-DALI;
- Bourne and Shindyalov, 1998 - CE.

Sequence order independent :

- Nussinov and Wolfson, 1991
 - Geometric Hashing.
- Mitchell et al., Koch et al. -
 - maximal clique in SSE correspondence graphs.

Previous Work - rigid alignment

Sequence order dependent :

- Taylor and Orengo, 1989 - SSAP;
- Holm and Sander, 1993 - FSSP-DALI;
- Bourne and Shindyalov, 1998 - CE.

Sequence order independent :

- Nussinov and Wolfson, 1991
 - Geometric Hashing.
- Mitchell et al., Koch et al. -
 - maximal clique in SSE correspondence graphs.

Previous Work – 3-D Curve Matching

–**Schwartz and Sharir, 1987,**
Matching of Noisy ‘Characteristic Curves’;

–**Kishon, Hastie and Wolfson, 1991,**
3-D Curve Matching Using Splines;

Previous Work - flexible alignment

Sequence order independent, flexible alignment (*hinge positions are input to the program*) :

–Fligelman, Nussinov, and Wolfson 2000

- Geometric Hashing, any number of hinges

–Verbitsky, Wolfson, and Nussinov 1999

- Pose Clustering, one hinge

–Rigoutsos, Platt, and Califano 1996

- Drug Databases

–Lengauer et al 1995

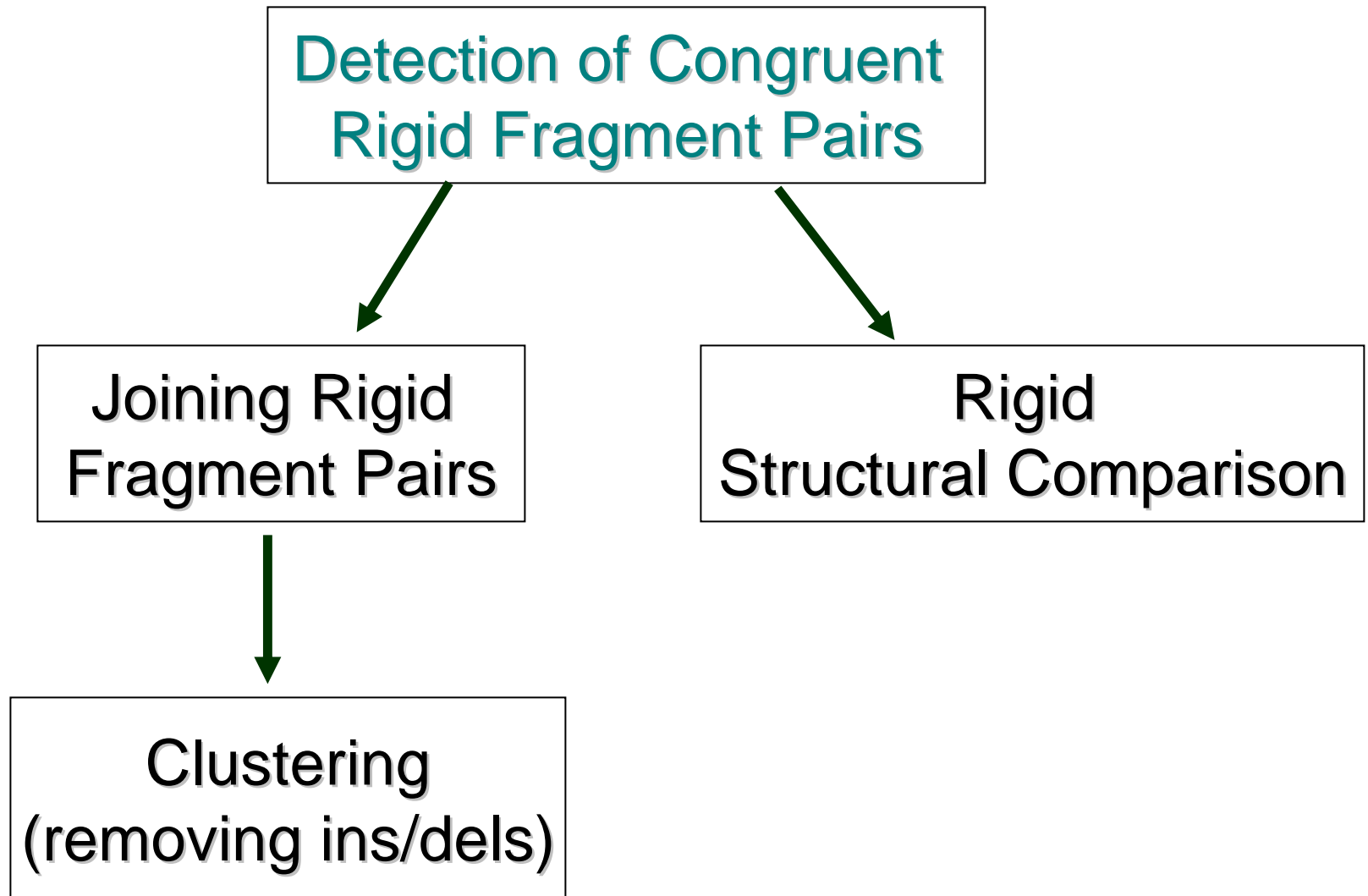
- FlexX - Docking

- FlexS - Structural Alignment

The Novel Algorithm

- FlexProt - flexible protein alignment :
 - sequence order dependent;
 - hypothesized hinge positions are detected automatically .

FlexProt Main Steps



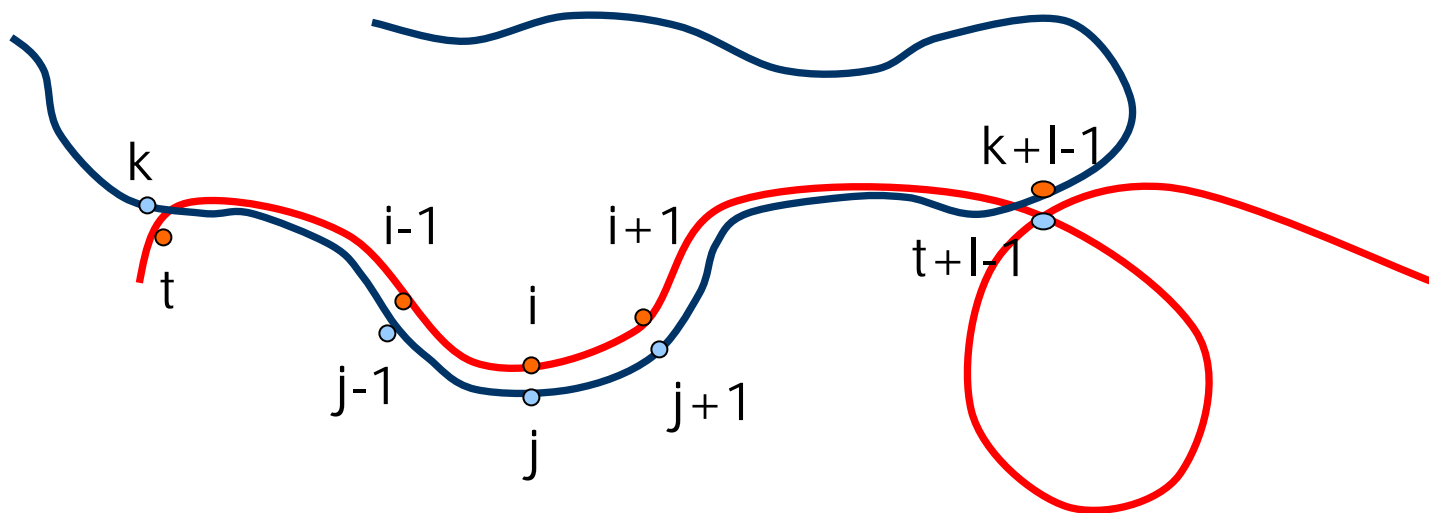
**Detection of Congruent
Rigid Fragment Pairs**

```
graph TD; A["Detection of Congruent Rigid Fragment Pairs"] --> B["Joining Rigid Fragment Pairs"]; B --> C["Clustering (removing ins/dels)"]
```

**Joining Rigid
Fragment Pairs**

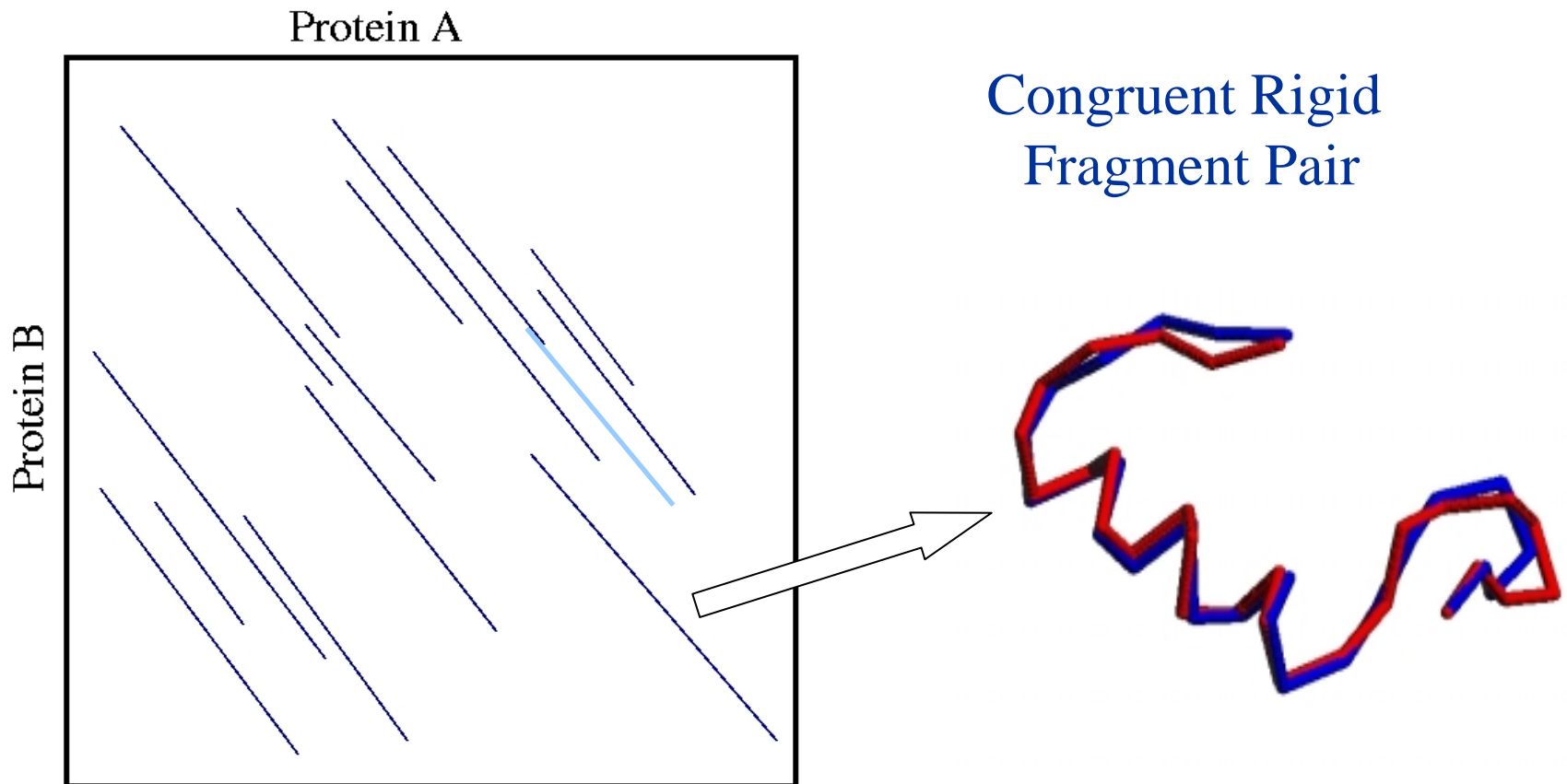
**Clustering
(removing ins/dels)**

Detection of Congruent Rigid Fragment Pairs



$$\begin{aligned}
 \mathit{Frag}_{kt}(l) = & \begin{matrix} \mathbf{V}_k & \dots & \mathbf{V}_i & \dots & \mathbf{V}_{k+l-1} \\ \mathbf{W}_t & \dots & \mathbf{W}_j & \dots & \mathbf{W}_{t+l-1} \end{matrix} \quad \text{RMSD}(\mathit{Frag}_{kt}(l)) < \varepsilon
 \end{aligned}$$

Structural Similarity Matrix



RMSD Computation

$$P = \begin{matrix} V_i & \dots & V_{i+1} \\ W_j & \dots & W_{j+1} \end{matrix}$$

$$Q = \begin{matrix} V_k & \dots & V_{k+m} \\ W_t & \dots & W_{t+m} \end{matrix}$$



$$P \cup Q$$

$$\text{RMSD}(P)$$

$$\text{RMSD}(Q)$$

$\text{RMSD}(P \cup Q)$ in $O(1)$ time

NOT $O(|P|+|Q|)$

RMSD Computation

$$\Delta = \min_T \sum_{i=1}^n |Tu_i - v_i|^2.$$

$$\begin{aligned} \Delta &= \sum_{i=1}^n |v_i|^2 - \frac{1}{n} \left| \sum_{i=1}^n v_i \right|^2 \\ &\quad + \sum_{i=1}^n |u_i|^2 \\ &\quad - 2tr((A^T A)^{1/2}) \end{aligned}$$

Where the 3×3 matrix A is defined by

$$\forall i, j = 1, 2, 3 \quad A_{ij} = \sum_{k=1}^n u_k^i v_k^j$$

Complexity of the First Step

$$T_1 = \sum_{F_i^1 F_j^2(l)} |\text{Time to compute } F_i^1 F_j^2(l)| \quad (5.1)$$

$$T_1 = \sum_{F_i^1 F_j^2(l)} O(l) \leq |\{(v_i, w_j)\}_{ij}| * O(n) = n^2 * O(n) \in O(n^3) \quad (5.2)$$

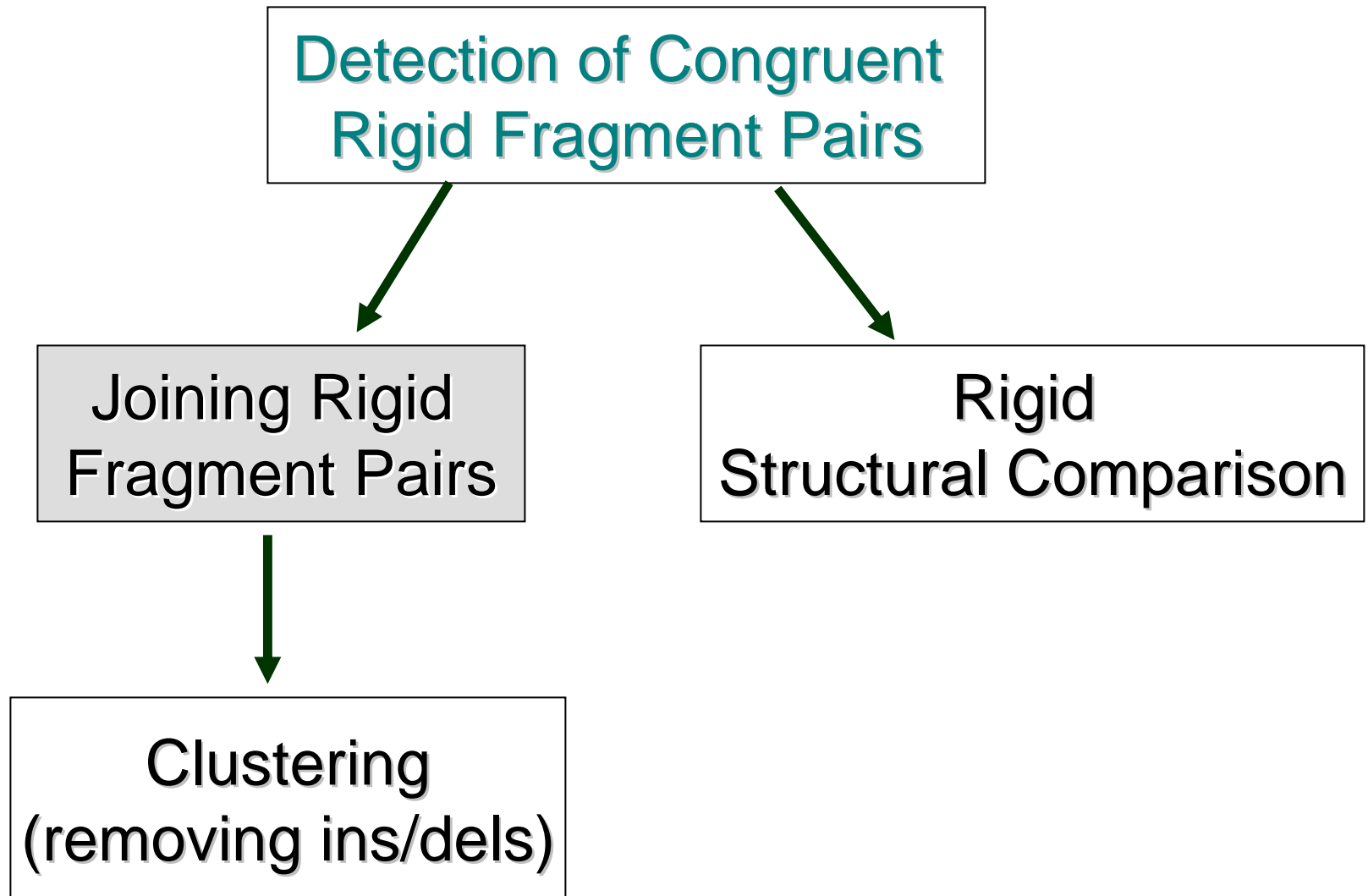
In Practice: $T_1 = \sum_{(v_i, w_j) \in F_k^1 F_t^2(l)} O(1) \in O(n^2)$

$$T_1 \in O(n^2)$$

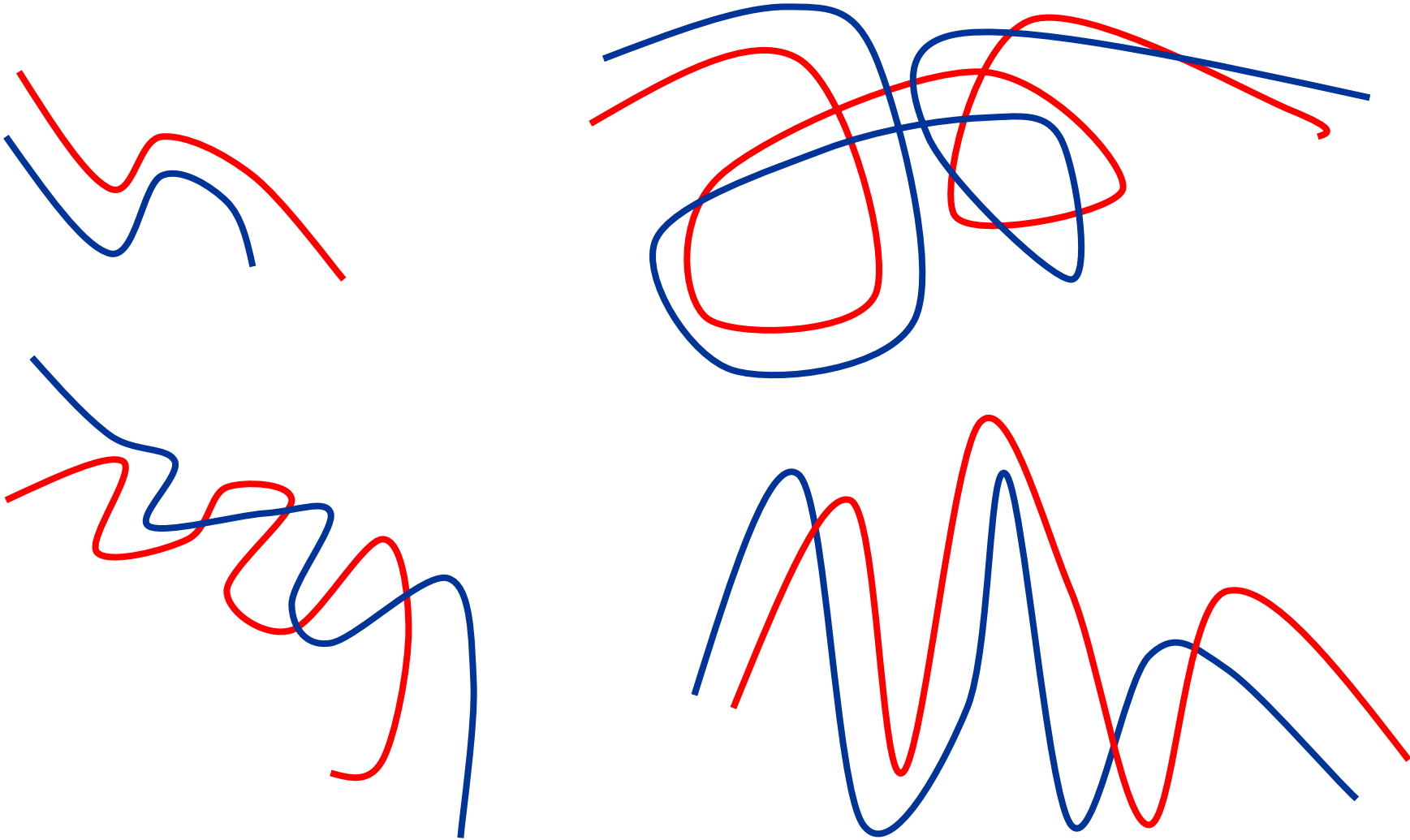
Number of Rigid Fragment Pairs:

$$|\{F_i^1 F_j^2(l)\}| \in O(n^2) \quad (5.4)$$

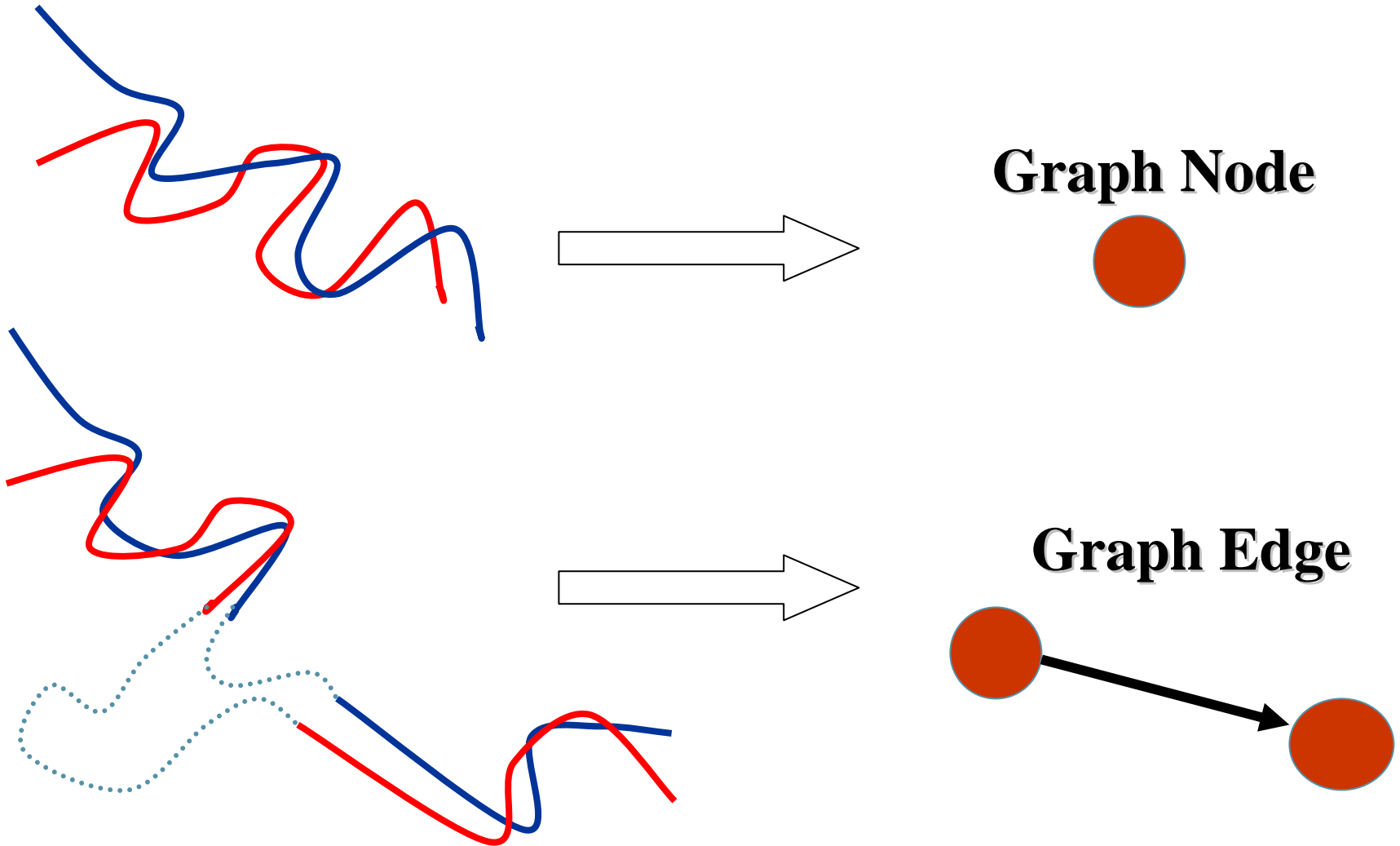
FlexProt Main Steps



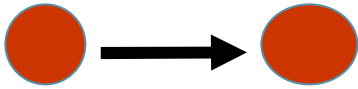
How to Join Rigid Fragment Pairs ?



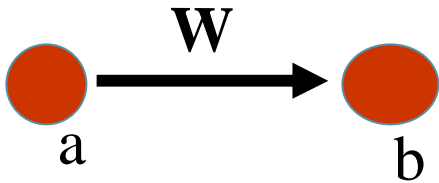
Graph Representation



Graph Representation



- **The fragments are in ascending order.**
- **The gaps (ins/dels) are limited.**
- **Allow some overlapping.**



$$w(e) = -((l + 1) - \lceil \Delta \rceil)^2 + \max(|Ins|, |Dels|) \\ + ||Ins| - |Dels||,$$

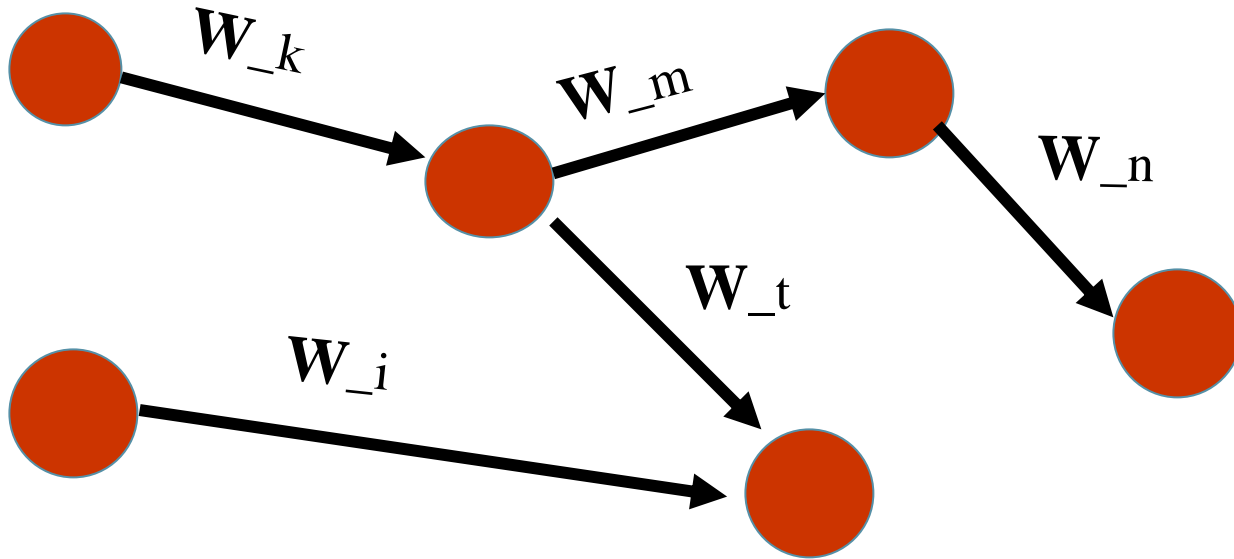
+ **Size of the rigid fragment pair (node b)**

- **Gaps (ins/dels)**

- **Overlapping**

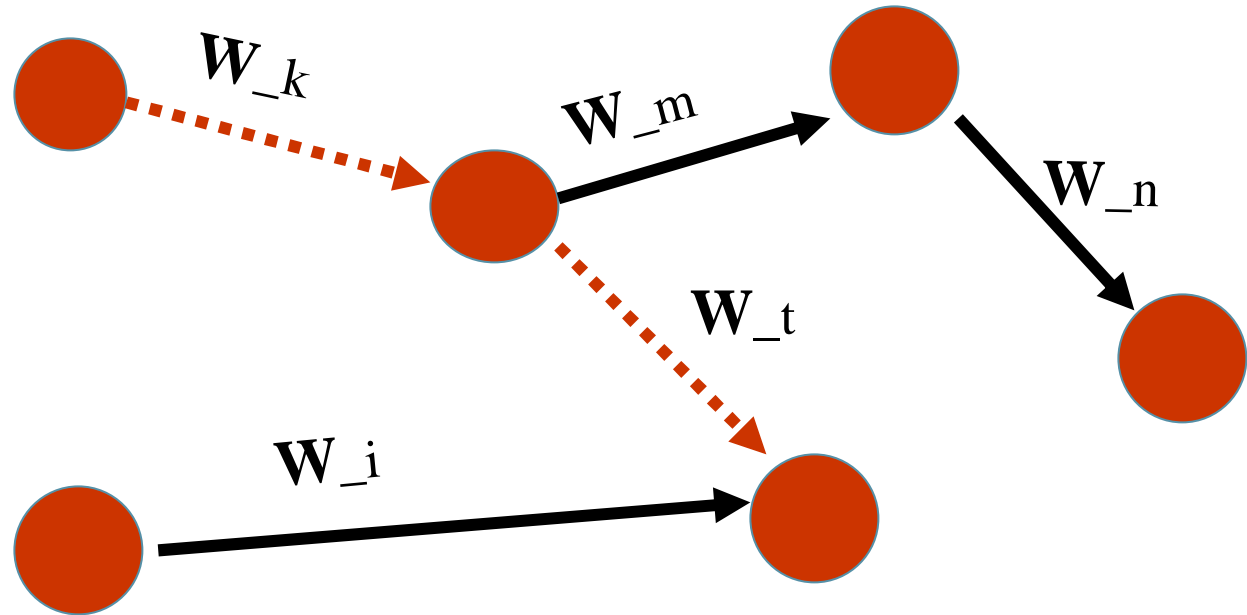
} Penalties

Graph Representation



- ***DAG*** (*directed acyclic graph*)

Optimal Solution



DAG -> in polynomial time, for each K , we can detect all weighted shortest paths which are exactly K nodes long.

-> all optimal flexible alignments are detected (containing $K-1$ hinges)

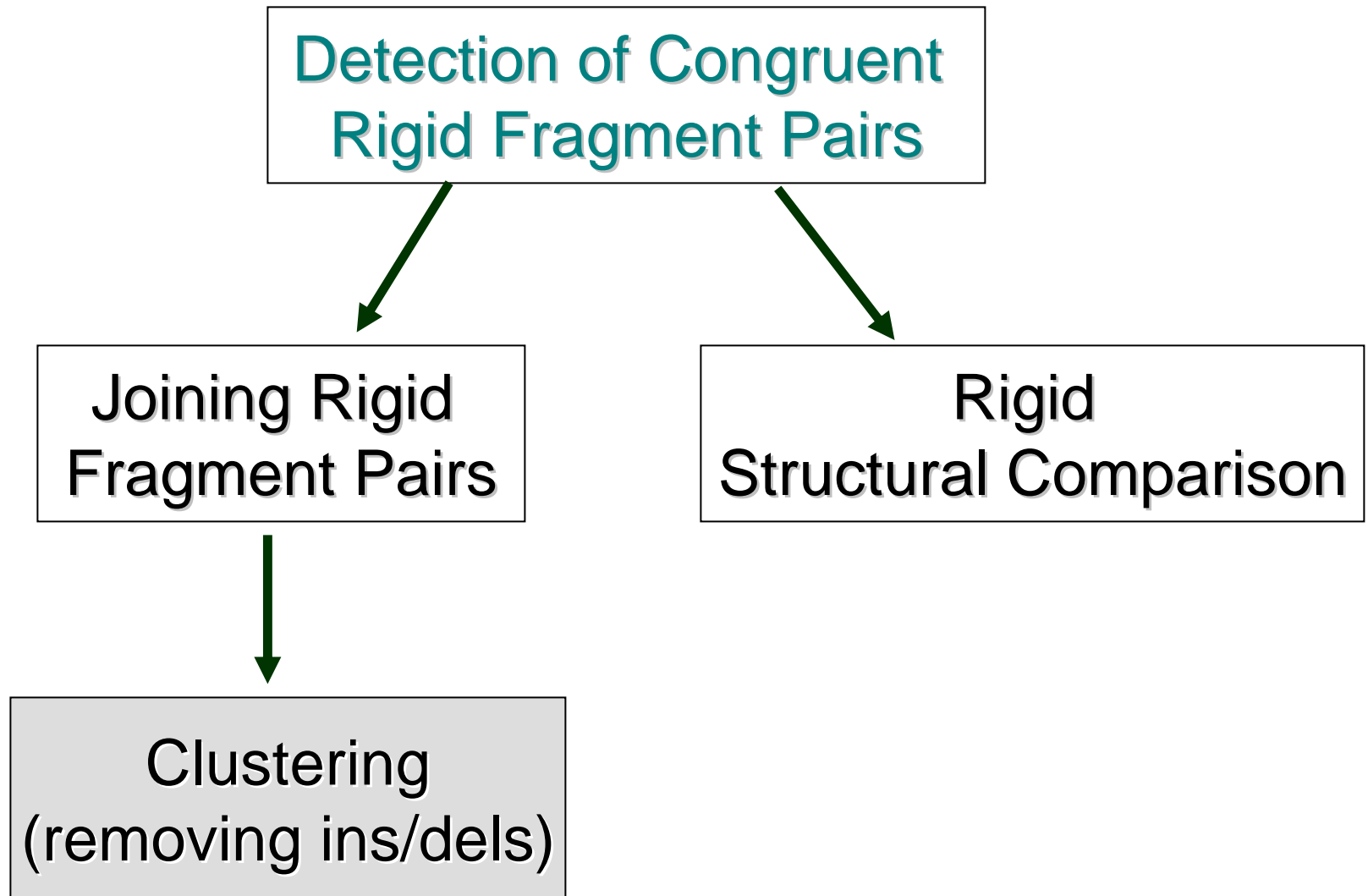
Complexity of the Second Step

$$T_2 = |V| + |E| + \text{MaxNumFlex}^2 * |V| + \quad (5.5)$$

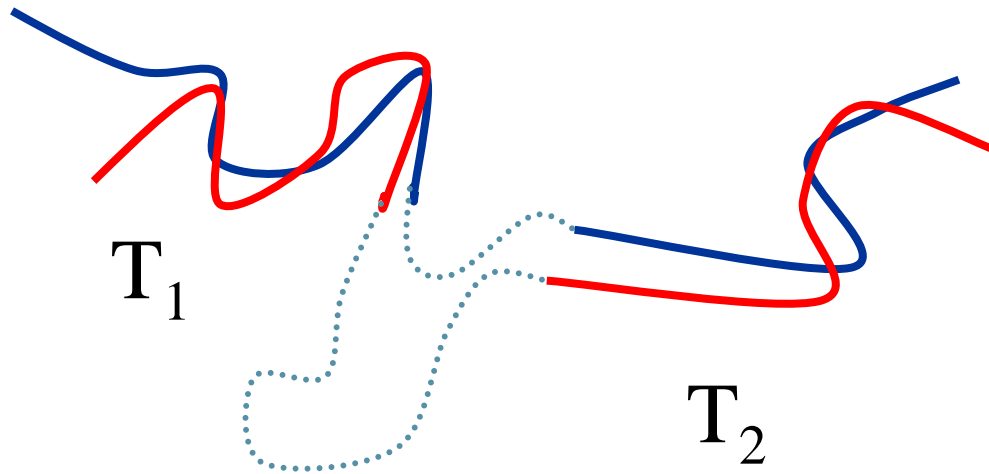
$$\text{MaxNumFlex} * |V| * \log(\text{MaxNumFlex} * |V|) \in O(|V|^2) \quad (5.6)$$

$$T_2 \in O(|V|^2) \in O(n^4) \quad (5.7)$$

FlexProt Main Steps



Clustering (removing ins/dels)



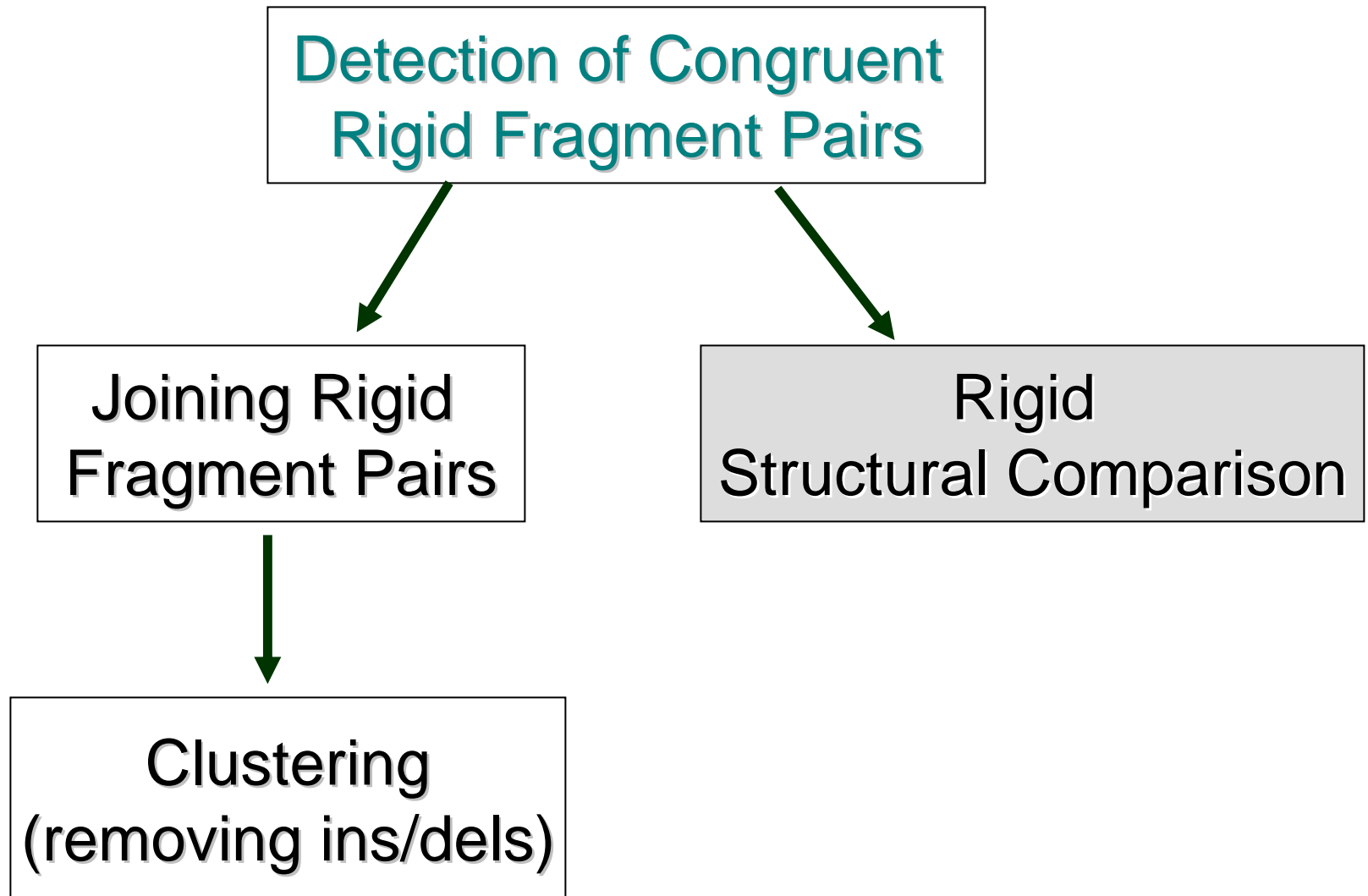
If joining two fragment pairs gives small RMSD ($T_1 \sim T_2$) then put them into one cluster.

Complexity of the Clustering Step

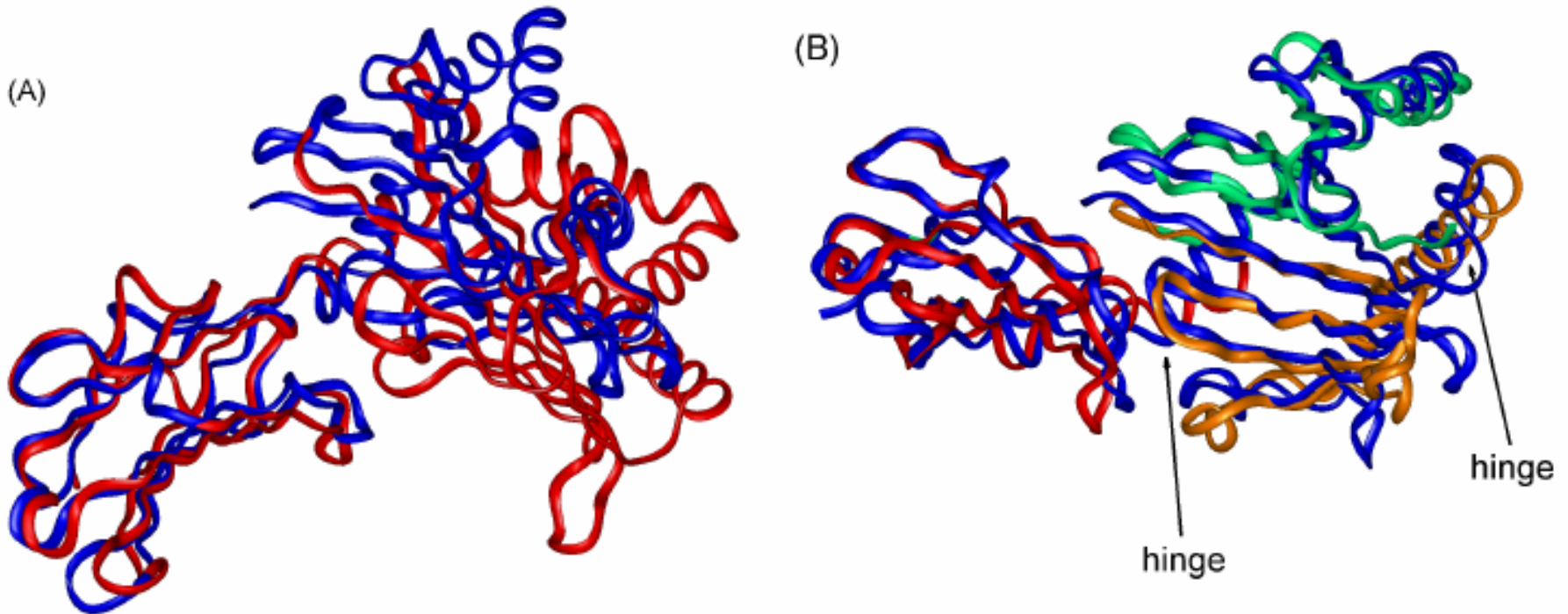
$T_3 = \text{Number Of Paths} * \text{Maximal Path Length}$

$$T_3 \in O(\text{MaxNumFlex}^2 * n^2) \quad (5.8)$$

FlexProt Main Steps



Experimental Results

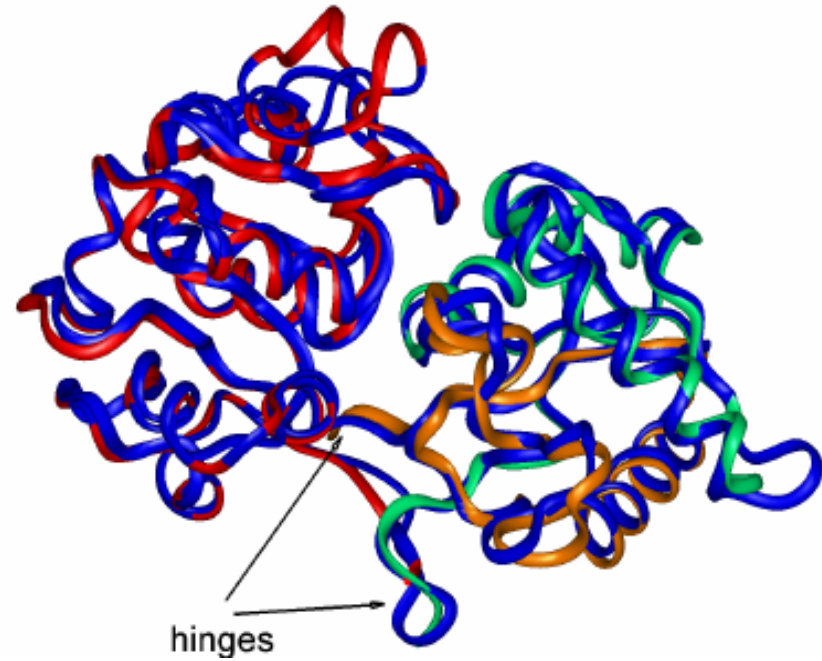


Experimental Results

(A)



(B)



FlexProt Web Site

<http://bioinfo3d.math.tau.ac.il/FlexProt>

Publications :

1. **M. Shatsky, Z.Y. Fligelman, R. Nussinov, H.J. Wolfson**, *Alignment of Flexible Protein Structures* Proc. of the 8'th International Conference on Intelligent Systems for Molecular Biology, San Diego, Ca., August 2000, pp. 329-343, (R. Altman et al.,ed.'s), AAAI Press, Menlo Park, California.
2. **M. Shatsky, R. Nussinov, H.J. Wolfson**, *Flexible Protein Alignment and Hinge Detection*, *Proteins* , 48(2), 242-256, (2002).

Publications (continued):

3. **G. Verbitsky, R. Nussinov, H.J. Wolfson**, *Flexible Structural Comparison Allowing Hinge Bending, Swivelling Motions*, *Proteins* , 34, 232-254, (1999).
4. **C. Lemmen, T. Lengauer, G. Klebe**, *FlexS: A Method for Fast Flexible Ligand Superposition*, *J. Med. Chem.* , 41, 4502-4520, (1998).