# Visualizing Projection Algorithms with Application to Protein Reconstruction

**Matthew K. Tam**

Joint work with Francisco Aragón Artacho and Jonathan Borwein

School of Mathematical and Physical Sciences
University of Newcastle, Australia




Challenges in 21st Century Experimental Mathematical Computation

21st-25th July 2014 at Brown University

# Introduction: Projection Methods

Projection methods are a family of iterative algorithms useful for solving the feasibility problem which asks:

$$\text{find } x \in C_1 \cap C_2 \subseteq \mathcal{H},$$

where $C_1$ and $C_2$ are constraint sets in a Hilbert space $\mathcal{H}$.

The focus of this talk is application of the Douglas–Rachford method as a heuristic for non-convex feasibility problems guided by convex theory.

---

Recall that a set $S$ is convex if, $\lambda x + (1-\lambda)y \in S$, $(\forall x, y \in S)(\forall \lambda \in [0,1])$.

# Introduction: Projection Methods

Projection methods are a family of iterative algorithms useful for solving the feasibility problem which asks:

$$\text{find } x \in C_1 \cap C_2 \subseteq \mathcal{H},$$

where $C_1$ and $C_2$ are constraint sets in a Hilbert space $\mathcal{H}$.

- At each stage, employ (nearest point) projections w.r.t. the individual constraint sets. The solution is obtained in the limit.

The focus of this talk is application of the Douglas–Rachford method as a heuristic for non-convex feasibility problems guided by convex theory.

---

Recall that a set $S$ is convex if, $\lambda x + (1-\lambda)y \in S$, $(\forall x, y \in S)(\forall \lambda \in [0,1])$.

# Introduction: Projection Methods

Projection methods are a family of iterative algorithms useful for solving the feasibility problem which asks:

$$\text{find } x \in C_1 \cap C_2 \subseteq \mathcal{H},$$

where $C_1$ and $C_2$ are constraint sets in a Hilbert space $\mathcal{H}$.

- At each stage, employ (nearest point) projections w.r.t. the individual constraint sets. The solution is obtained in the limit.
- For (closed) convex constraint sets, behavior is *fairly* well understood – the methods can be analyzed using non-expansivity properties of the convex projection operators.

The focus of this talk is application of the Douglas–Rachford method as a heuristic for non-convex feasibility problems guided by convex theory.

---

Recall that a set $S$ is convex if, $\lambda x + (1-\lambda)y \in S$, $(\forall x, y \in S)(\forall \lambda \in [0,1])$.

# Introduction: Projection Methods

Projection methods are a family of iterative algorithms useful for solving the feasibility problem which asks:

$$\text{find } x \in C_1 \cap C_2 \subseteq \mathcal{H},$$

where $C_1$ and $C_2$ are constraint sets in a Hilbert space $\mathcal{H}$.

- At each stage, employ (nearest point) projections w.r.t. the individual constraint sets. The solution is obtained in the limit.
- For (closed) convex constraint sets, behavior is *fairly* well understood – the methods can be analyzed using non-expansivity properties of the convex projection operators.
- When one or more of the constraint sets are non-convex, theory is largely unknown. However, one particular projection method, the Douglas–Rachford method, has been (experimentally) observed to successfully solve a large range of non-convex problems.

The focus of this talk is application of the Douglas–Rachford method as a heuristic for non-convex feasibility problems guided by convex theory.

---

Recall that a set $S$ is convex if, $\lambda x + (1-\lambda)y \in S$, $(\forall x, y \in S)(\forall \lambda \in [0,1])$.

# Introduction: Projection Methods

Projection methods are a family of iterative algorithms useful for solving the feasibility problem which asks:

$$\text{find } x \in C_1 \cap C_2 \subseteq \mathcal{H},$$

where $C_1$ and $C_2$ are constraint sets in a Hilbert space $\mathcal{H}$.

- At each stage, employ (nearest point) projections w.r.t. the individual constraint sets. The solution is obtained in the limit.
- For (closed) convex constraint sets, behavior is *fairly* well understood – the methods can be analyzed using non-expansivity properties of the convex projection operators.
- When one or more of the constraint sets are non-convex, theory is largely unknown. However, one particular projection method, the Douglas–Rachford method, has been (experimentally) observed to successfully solve a large range of non-convex problems. Examples:
  - Solving Sudoku and nonogram puzzles, 8-queens and generalizations, enumerating Hadamard matrices, phase retrieval & ptychography, ...

The focus of this talk is application of the Douglas–Rachford method as a heuristic for non-convex feasibility problems guided by convex theory.

---

Recall that a set $S$ is convex if, $\lambda x + (1-\lambda)y \in S$, $(\forall x, y \in S)(\forall \lambda \in [0,1])$.

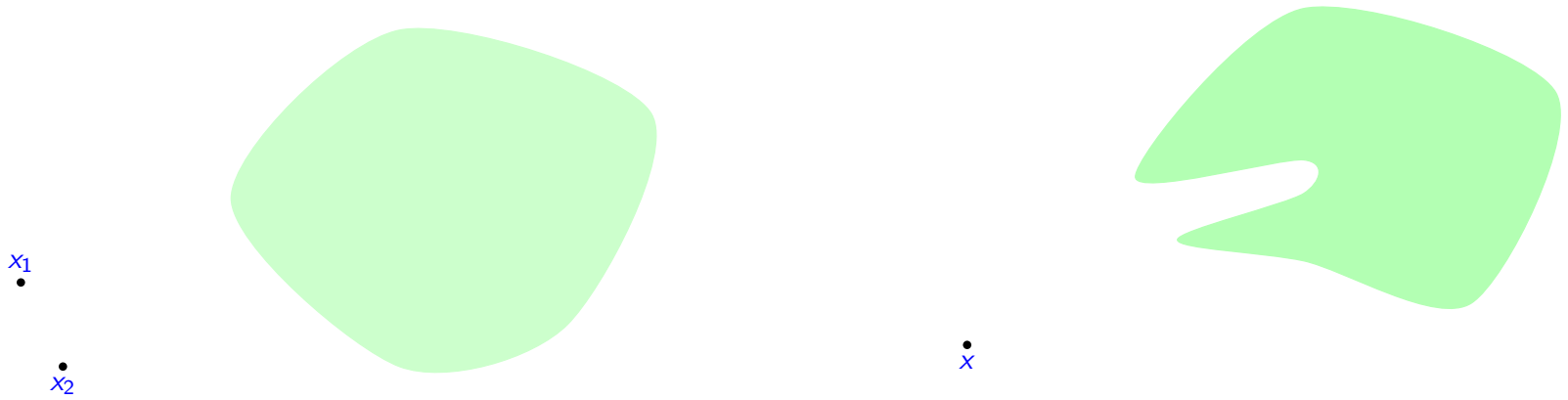
# Introduction: Variational Tools

Let $S \subseteq \mathcal{H}$. The (nearest point) projection onto $S$ is the (set-valued) mapping,

$$P_S x := \operatorname*{arg\,min}_{s \in S} \|s - x\|.$$

The reflection w.r.t. $S$ is the (set-valued) mapping,

$$R_S := 2P_S - I.$$



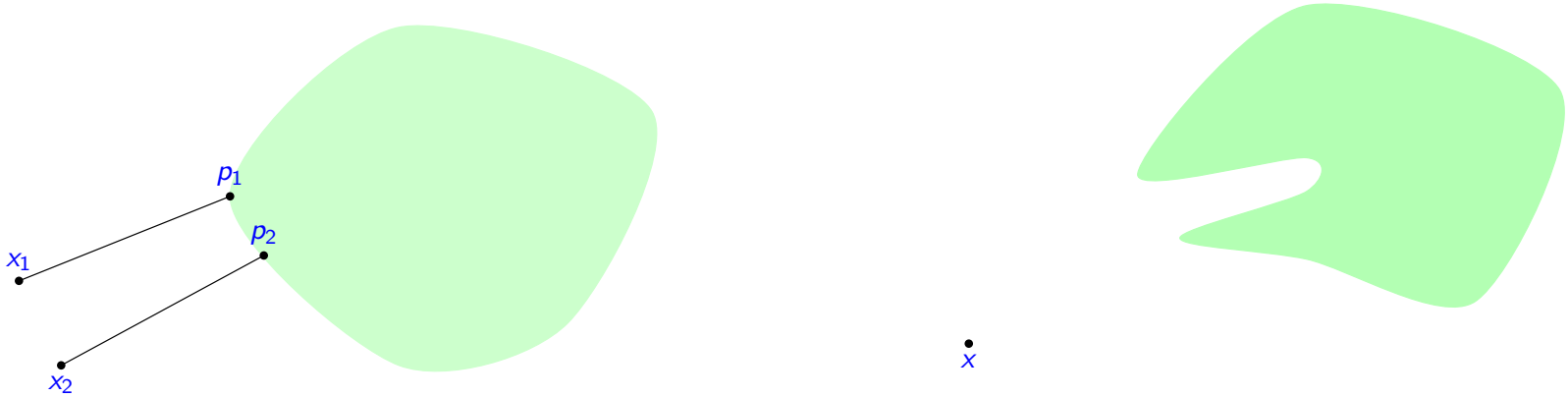
# Introduction: Variational Tools

Let $S \subseteq \mathcal{H}$. The (nearest point) projection onto $S$ is the (set-valued) mapping,

$$P_S x := \operatorname*{arg\,min}_{s \in S} \|s - x\|.$$

The reflection w.r.t. $S$ is the (set-valued) mapping,

$$R_S := 2P_S - I.$$



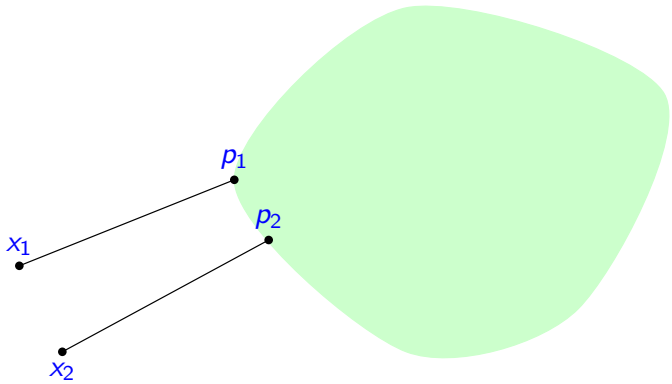
# Introduction: Variational Tools

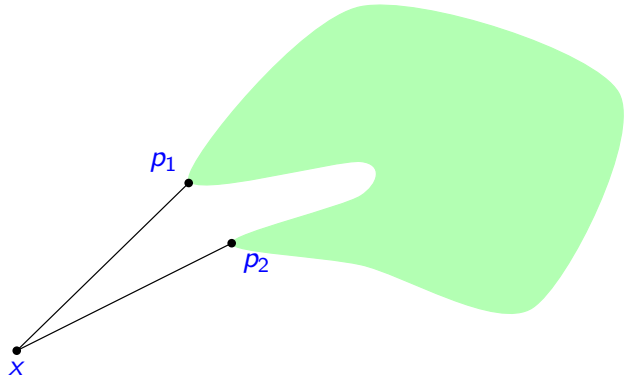
Let $S \subseteq \mathcal{H}$. The (nearest point) projection onto $S$ is the (set-valued) mapping,

$$P_S x := \operatorname*{arg\,min}_{s \in S} \|s - x\|.$$

The reflection w.r.t. $S$ is the (set-valued) mapping,

$$R_S := 2P_S - I.$$



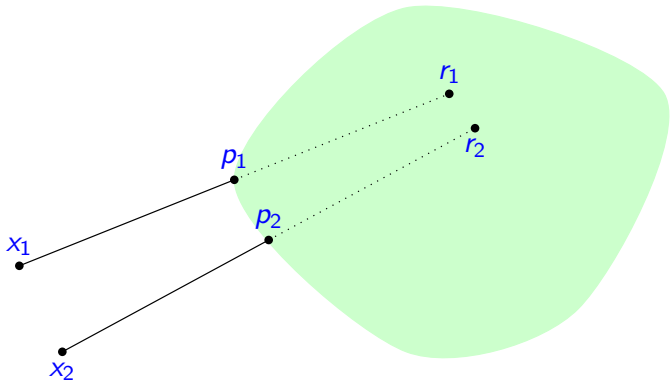
# Introduction: Variational Tools

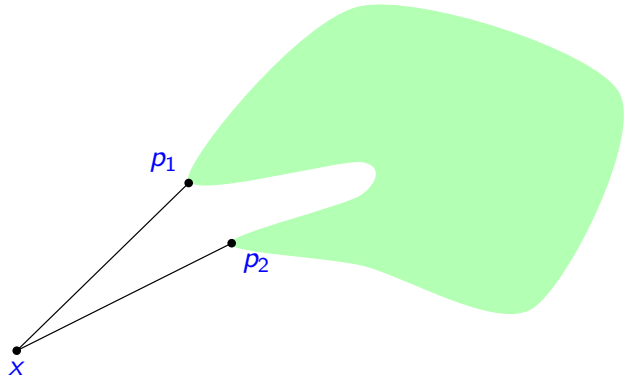
Let $S \subseteq \mathcal{H}$. The (nearest point) projection onto $S$ is the (set-valued) mapping,

$$P_S x := \operatorname*{arg\,min}_{s \in S} \|s - x\|.$$

The reflection w.r.t. $S$ is the (set-valued) mapping,

$$R_S := 2P_S - I.$$



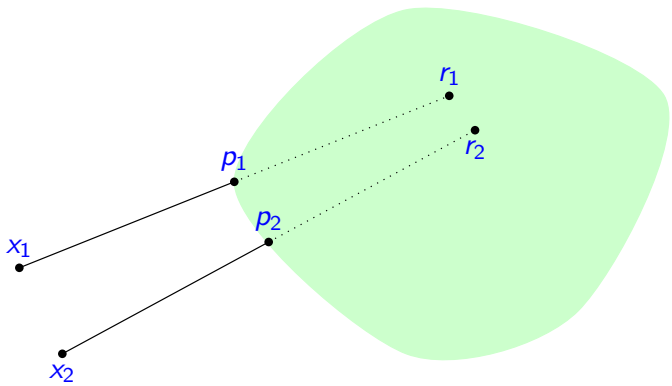
# Introduction: Variational Tools

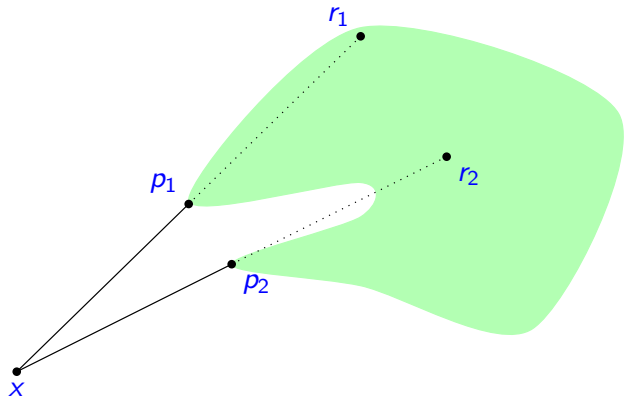
Let $S \subseteq \mathcal{H}$. The (nearest point) projection onto $S$ is the (set-valued) mapping,

$$P_S x := \operatorname*{arg\,min}_{s \in S} \|s - x\|.$$

The reflection w.r.t. $S$ is the (set-valued) mapping,

$$R_S := 2P_S - I.$$



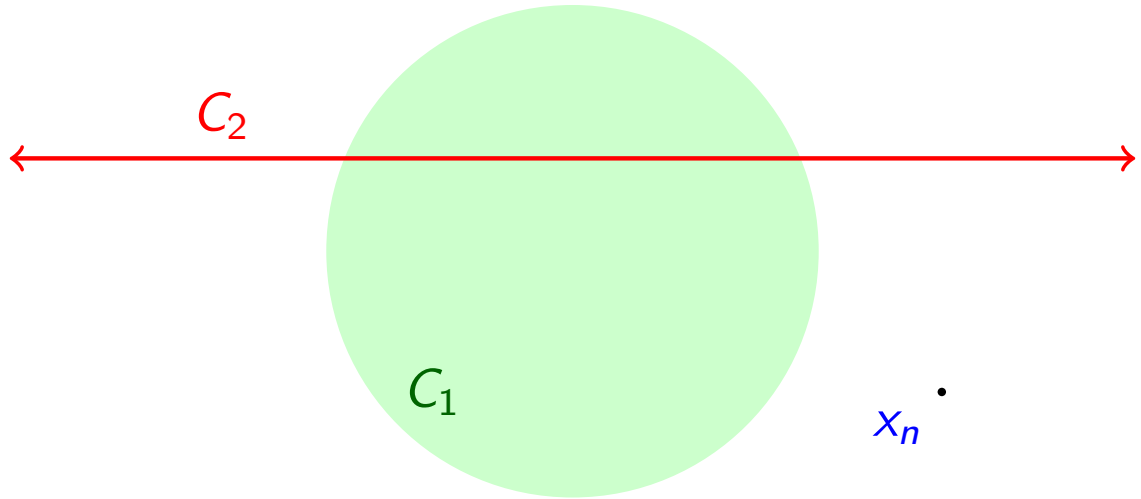



# The Douglas–Rachford Algorithm

Given an initial point $x_0 \in \mathcal{H}$, the Douglas–Rachford method is the fixed-point iteration given by

$$x_{n+1} = T_{C_1,C_2} x_n \quad \text{where} \quad T_{C_1,C_2} := \frac{Id + R_{C_2} R_{C_1}}{2}.$$

- If $x$ is a fixed point of $T_{C_1,C_2}$ then $P_{C_1} x \in C_1 \cap C_2$.



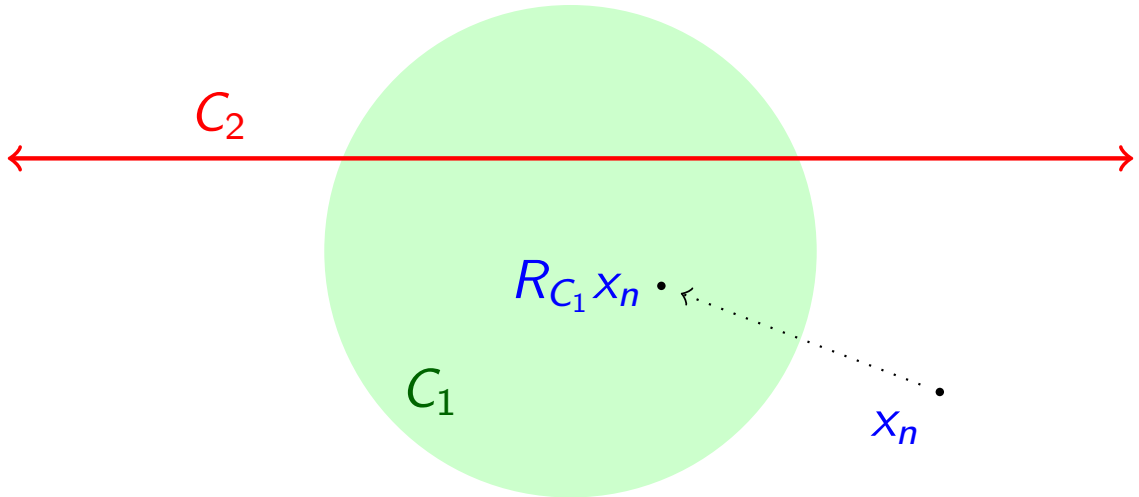

$$C_1 = \{x \in \mathcal{H} : \|x\| \leq 1\}, \quad C_2 = \{x \in \mathcal{H} : \langle a, x \rangle = b\}.$$

# The Douglas–Rachford Algorithm

Given an initial point $x_0 \in \mathcal{H}$, the Douglas–Rachford method is the fixed-point iteration given by

$$x_{n+1} = T_{C_1,C_2} x_n \quad \text{where} \quad T_{C_1,C_2} := \frac{Id + R_{C_2} R_{C_1}}{2}.$$

- If $x$ is a fixed point of $T_{C_1,C_2}$ then $P_{C_1} x \in C_1 \cap C_2$.



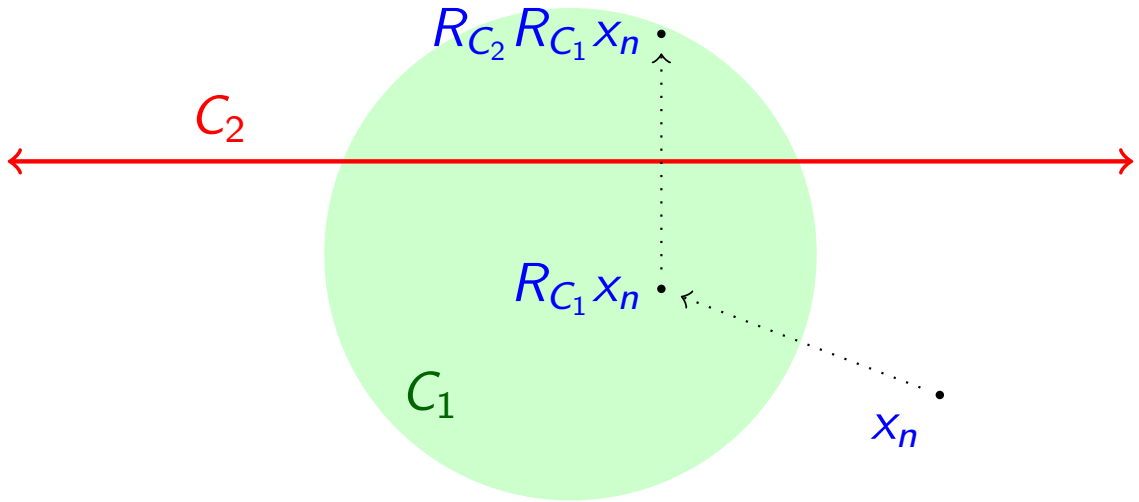

$$C_1 = \{x \in \mathcal{H} : \|x\| \leq 1\}, \quad C_2 = \{x \in \mathcal{H} : \langle a, x \rangle = b\}.$$

# The Douglas–Rachford Algorithm

Given an initial point $x_0 \in \mathcal{H}$, the Douglas–Rachford method is the fixed-point iteration given by

$$x_{n+1} = T_{C_1,C_2} x_n \quad \text{where} \quad T_{C_1,C_2} := \frac{Id + R_{C_2} R_{C_1}}{2}.$$

- If $x$ is a fixed point of $T_{C_1,C_2}$ then $P_{C_1} x \in C_1 \cap C_2$.



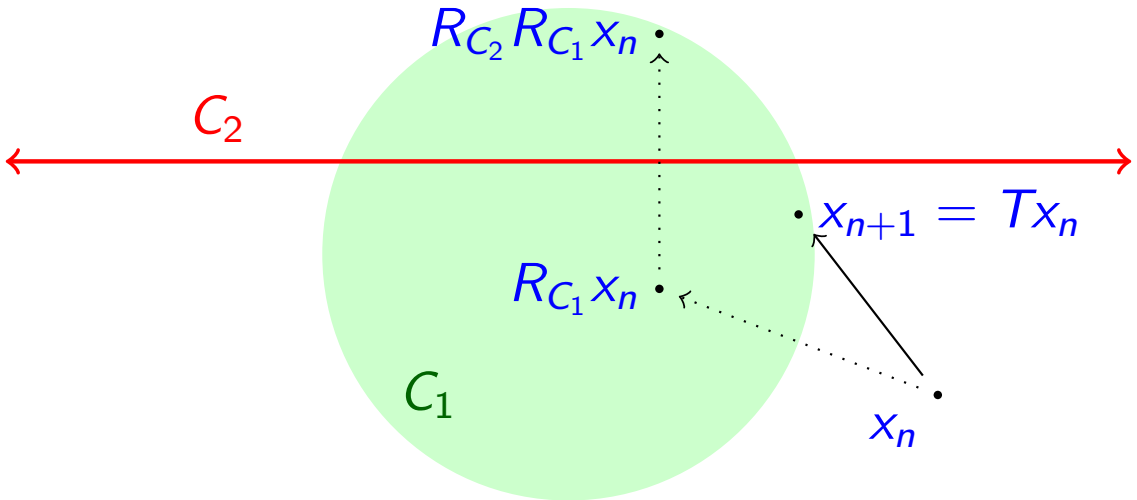

$$C_1 = \{x \in \mathcal{H} : \|x\| \leq 1\}, \quad C_2 = \{x \in \mathcal{H} : \langle a, x \rangle = b\}.$$

# The Douglas–Rachford Algorithm

Given an initial point $x_0 \in \mathcal{H}$, the Douglas–Rachford method is the fixed-point iteration given by

$$x_{n+1} = T_{C_1,C_2} x_n \quad \text{where} \quad T_{C_1,C_2} := \frac{Id + R_{C_2} R_{C_1}}{2}.$$

- If $x$ is a fixed point of $T_{C_1,C_2}$ then $P_{C_1} x \in C_1 \cap C_2$.



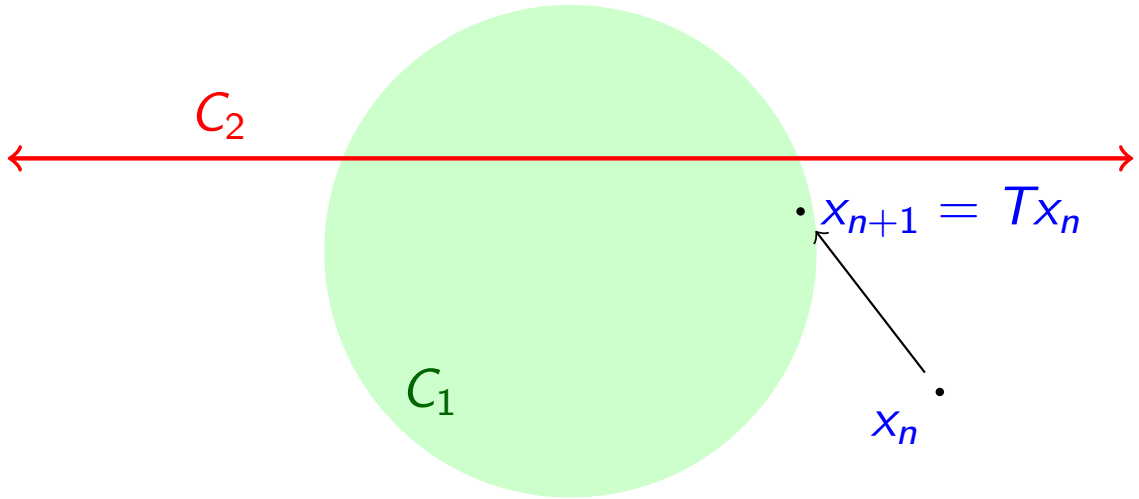

$$C_1 = \{x \in \mathcal{H} : \|x\| \leq 1\}, \quad C_2 = \{x \in \mathcal{H} : \langle a, x \rangle = b\}.$$

# The Douglas–Rachford Algorithm

Given an initial point $x_0 \in \mathcal{H}$, the Douglas–Rachford method is the fixed-point iteration given by

$$x_{n+1} = T_{C_1,C_2} x_n \quad \text{where} \quad T_{C_1,C_2} := \frac{Id + R_{C_2} R_{C_1}}{2}.$$

- If $x$ is a fixed point of $T_{C_1,C_2}$ then $P_{C_1} x \in C_1 \cap C_2$.




$$C_1 = \{x \in \mathcal{H} : \|x\| \leq 1\}, \quad C_2 = \{x \in \mathcal{H} : \langle a, x \rangle = b\}.$$

# The Douglas–Rachford Algorithm

- First studied by Douglas & Rachford (1956) in connection with heat conduction problems, and later by Lions & Mercier (1979) for finding a zero in the sum of two maximal monotone operators.

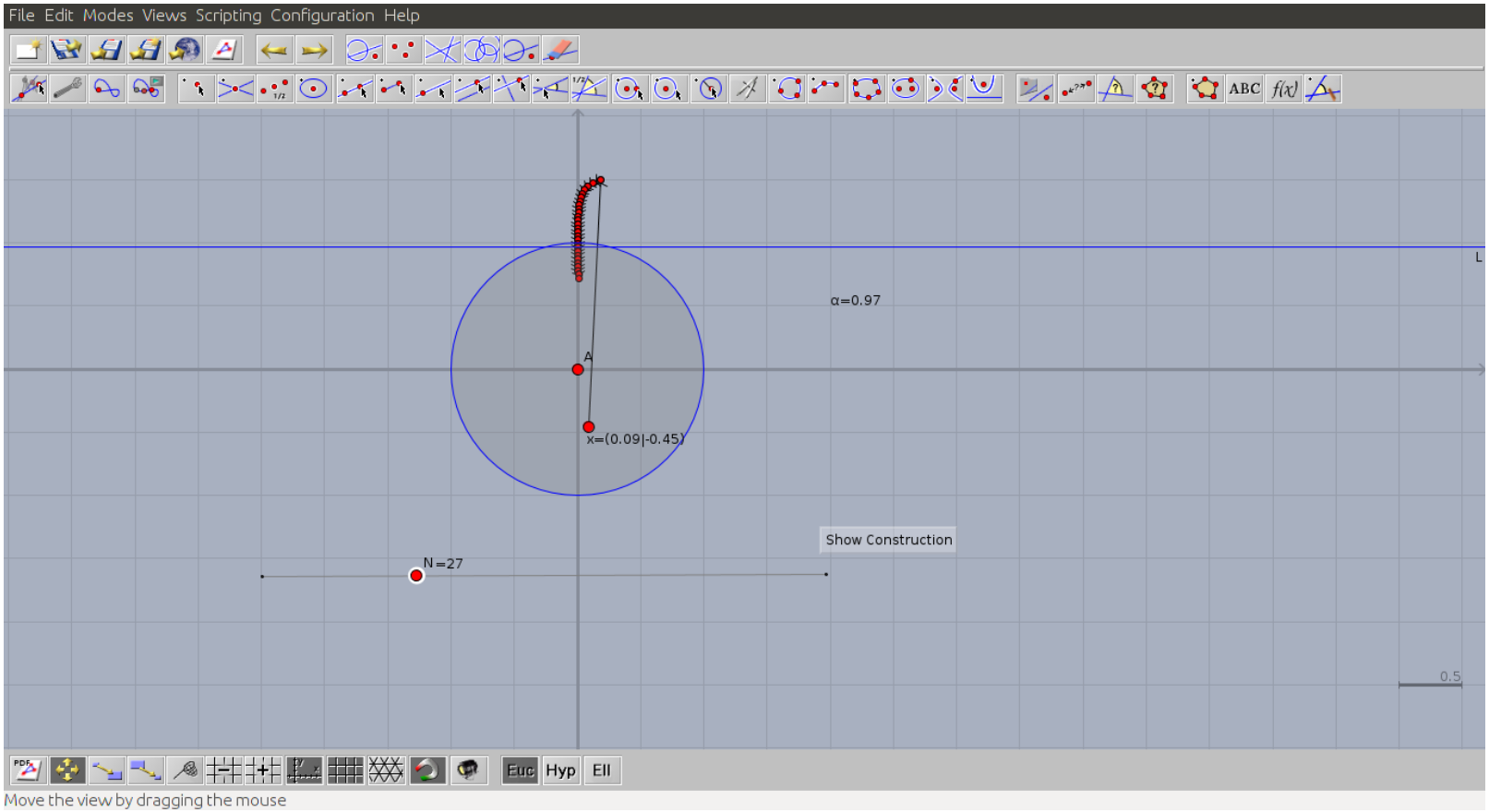
**Theorem (Basic behaviour of the Douglas–Rachford method)**

Suppose $C_1, C_2$ are closed convex subsets of a finite dimensional Hilbert space $\mathcal{H}$. For any $x_0 \in \mathcal{H}$, define $x_{n+1} = T_{C_1,C_2} x_n$.

1. If $C_1 \cap C_2 \neq \emptyset$, then $x_n \to x$ such that $P_{C_1} x \in C_1 \cap C_2$.
2. If $C_1 \cap C_2 = \emptyset$, then $\|x_n\| \to +\infty$.

- It is important to monitor the shadow sequence $(P_{C_1} x_n)_{n=1}^{\infty}$, not just the iterates $(x_n)_{n=1}^{\infty}$.

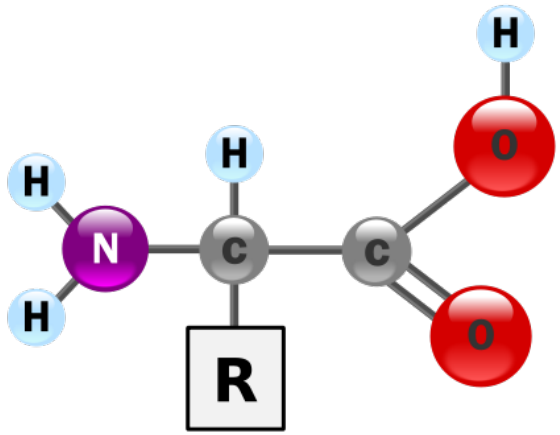
## *Cinderella*: Interactive Geometry




http://carma.newcastle.edu.au/jon/reflection.html

# Protein Confirmation Determination and EDMs

Proteins are large biomolecules comprising of multiple amino acid chains.




Generic amino acid

Myoglobin

They participate in virtually every cellular process, and knowledge of structural conformation gives insights into the mechanisms by which they perform.

# Protein Confirmation Determination and EDMs

One technique that can be used to determine conformation is nuclear magnetic resonance (NMR) spectroscopy. However, NMR is only able to resolve short inter-atomic distances (*i.e.*, $< 6\text{Å}$).

# Protein Confirmation Determination and EDMs

One technique that can be used to determine conformation is nuclear magnetic resonance (NMR) spectroscopy. However, NMR is only able to resolve short inter-atomic distances (*i.e.*, $< 6$Å). For 1PTQ (404 atoms) this corresponds to $< 8\%$ of the total inter-atomic distances.

# Protein Confirmation Determination and EDMs

One technique that can be used to determine conformation is nuclear magnetic resonance (NMR) spectroscopy. However, NMR is only able to resolve short inter-atomic distances (*i.e.*, $< 6$Å). For 1PTQ (404 atoms) this corresponds to $< 8\%$ of the total inter-atomic distances.

We say $D = (D_{ij}) \in \mathbb{R}^{m \times m}$ is a Euclidean distance matrix (EDM) if there exists points $p_1, \ldots, p_m \in \mathbb{R}^q$ such that

$$D_{ij} = \|p_i - p_j\|^2.$$

When this holds for points in $\mathbb{R}^q$, we say that $D$ is embeddable in $\mathbb{R}^q$.

# Protein Confirmation Determination and EDMs

One technique that can be used to determine conformation is nuclear magnetic resonance (NMR) spectroscopy. However, NMR is only able to resolve short inter-atomic distances (*i.e.*, $< 6$Å). For 1PTQ (404 atoms) this corresponds to $< 8\%$ of the total inter-atomic distances.

We say $D = (D_{ij}) \in \mathbb{R}^{m \times m}$ is a Euclidean distance matrix (EDM) if there exists points $p_1, \ldots, p_m \in \mathbb{R}^q$ such that

$$D_{ij} = \|p_i - p_j\|^2.$$

When this holds for points in $\mathbb{R}^q$, we say that $D$ is embeddable in $\mathbb{R}^q$.

We formulate protein reconstruction as a matrix completion problem:

> *Find a member from a given family of matrices,*
> *knowing only a subset of its entries.*

Find a EDM, embeddable in $\mathbb{R}^3$, knowing only short inter-atomic distances.

# A Feasibility Problem Formulation

Denote by $Q$ the Householder matrix defined by

$$Q := I - \frac{2vv^T}{v^T v}, \text{where } v = \left[1, 1, \ldots, 1, 1 + \sqrt{m}\right]^T \in \mathbb{R}^m.$$

**Theorem (Hayden–Wells 1988)**

A nonnegative, symmetric, hollow matrix $X$, is a EDM iff $\widehat{X} \in \mathbb{R}^{(m-1)\times(m-1)}$ in

$$Q(-X)Q = \begin{bmatrix} \widehat{X} & d \\ d^T & \delta \end{bmatrix} \qquad (*)$$

is positive semi-definite (PSD). In this case, $X$ is embeddable in $\mathbb{R}^q$ where $q = \text{rank}(\widehat{X}) \leq m - 1$ but not in $\mathbb{R}^{q-1}$.

Let $D$ denote the partial EDM (obtained from NMR), and $\Omega \subset \mathbb{N} \times \mathbb{N}$ the set of indices for known entries. In light of the above characterization, the protein reconstruction problem is the feasibility problem with constraints:

$$C_1 = \{X \in \mathbb{R}^{m\times m} : X \geq 0, X_{ij} = D_{ij} \text{ for } (i,j) \in \Omega\},$$

$$C_2 = \{X \in \mathbb{R}^{m\times m} : \widehat{X} \text{ in } (*) \text{ is PSD with } \text{rank}\,\widehat{X} \leq 3\}.$$

# A Feasibility Problem Formulation

Recall the constraint sets:

$$C_1 = \{X \in \mathbb{R}^{m\times m} : X \geq 0, X_{ij} = D_{ij} \text{ for } (i,j) \in \Omega\},$$

$$C_2 = \{X \in \mathbb{R}^{m\times m} : \widehat{X} \text{ in } (*) \text{ is PSD with } \operatorname{rank} \widehat{X} \leq 3\}.$$

# A Feasibility Problem Formulation

Recall the constraint sets:

$$C_1 = \{X \in \mathbb{R}^{m\times m} : X \geq 0, X_{ij} = D_{ij} \text{ for } (i,j) \in \Omega\},$$
$$C_2 = \{X \in \mathbb{R}^{m\times m} : \widehat{X} \text{ in } (*) \text{ is PSD with } \operatorname{rank} \widehat{X} \leq 3\}.$$

Now,

- $C_1$ is a convex set (intersection of cone and affine subspace).
- $C_2$ is convex iff $m \leq 2$ (in which case $C_2 = \mathbb{R}^{m\times m}$).

For interesting problems, $C_2$ is **never convex**.

# Computing Projections and Reflections

Recall the constraint sets:

$$C_1 = \{X \in \mathbb{R}^{m\times m} : X \geq 0, X_{ij} = D_{ij} \text{ for } (i,j) \in \Omega\},$$

$$C_2 = \{X \in \mathbb{R}^{m\times m} : \widehat{X} \text{ in } (*) \text{ is PSD with } \operatorname{rank} \widehat{X} \leq 3\}.$$

# Computing Projections and Reflections

Recall the constraint sets:

$$C_1 = \{X \in \mathbb{R}^{m\times m} : X \geq 0, X_{ij} = D_{ij} \text{ for } (i,j) \in \Omega\},$$
$$C_2 = \{X \in \mathbb{R}^{m\times m} : \widehat{X} \text{ in } (*) \text{ is PSD with } \operatorname{rank} \widehat{X} \leq 3\}.$$

The projection onto $C_1$ is given (point-wise) by

$$P_{C_1}(X)_{ij} = \begin{cases} D_{ij} & \text{if } (i,j) \in \Omega, \\ \max\{0, X_{ij}\} & \text{otherwise.} \end{cases}$$

# Computing Projections and Reflections

Recall the constraint sets:

$$C_1 = \{X \in \mathbb{R}^{m\times m} : X \geq 0, X_{ij} = D_{ij} \text{ for } (i,j) \in \Omega\},$$
$$C_2 = \{X \in \mathbb{R}^{m\times m} : \widehat{X} \text{ in } (*) \text{ is PSD with } \operatorname{rank} \widehat{X} \leq 3\}.$$

The projection onto $C_1$ is given (point-wise) by

$$P_{C_1}(X)_{ij} = \begin{cases} D_{ij} & \text{if } (i,j) \in \Omega, \\ \max\{0, X_{ij}\} & \text{otherwise.} \end{cases}$$

The projection onto $C_2$ is the set

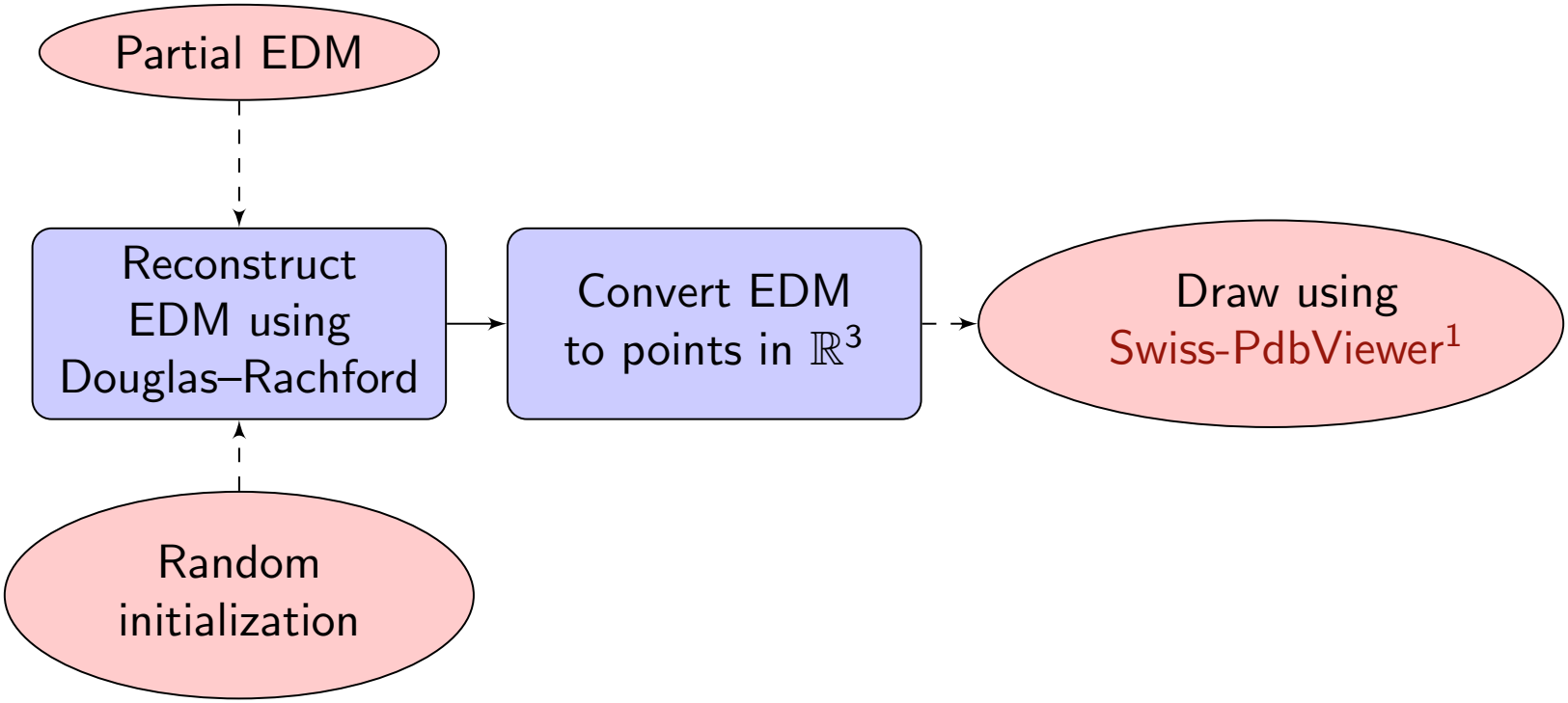
$$P_{C_2}(X) = \left\{ -Q \begin{bmatrix} \widehat{Y} & d \\ d^T & \delta \end{bmatrix} Q : Q(-X)Q = \begin{bmatrix} \widehat{X} & d \\ d^T & \delta \end{bmatrix}, \begin{array}{l} \widehat{X} \in \mathbb{R}^{(m-1)\times(m-1)}, \\ d \in \mathbb{R}^{m-1}, \delta \in \mathbb{R}, \end{array} \widehat{Y} \in P_S \widehat{X} \right\},$$

where $S$ is the set of PSD matrices of rank 3 or less.

- One method to compute $P_S$ is using the eigen-decomposition of $\widehat{X}$.

# Numerical and Visual Experiments

The reconstruction approach is as follows:



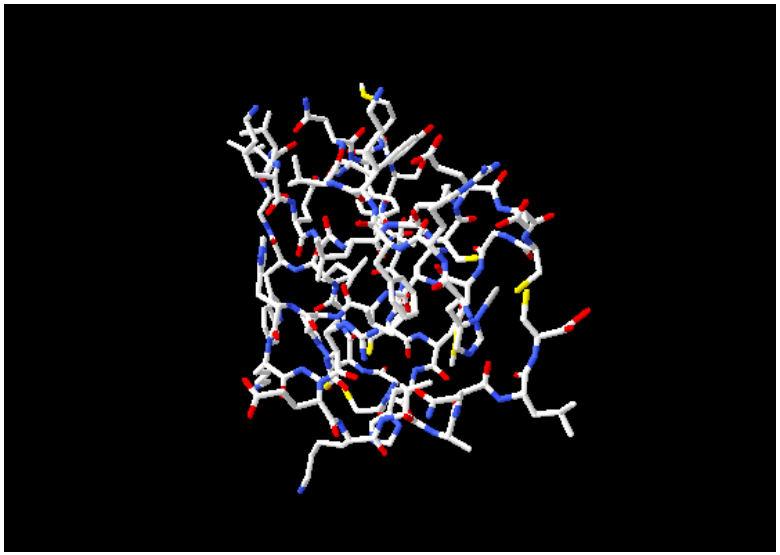

[1] http://spdbv.vital-it.ch/

# Experiment 1: Does the Douglas–Rachford Method Work?

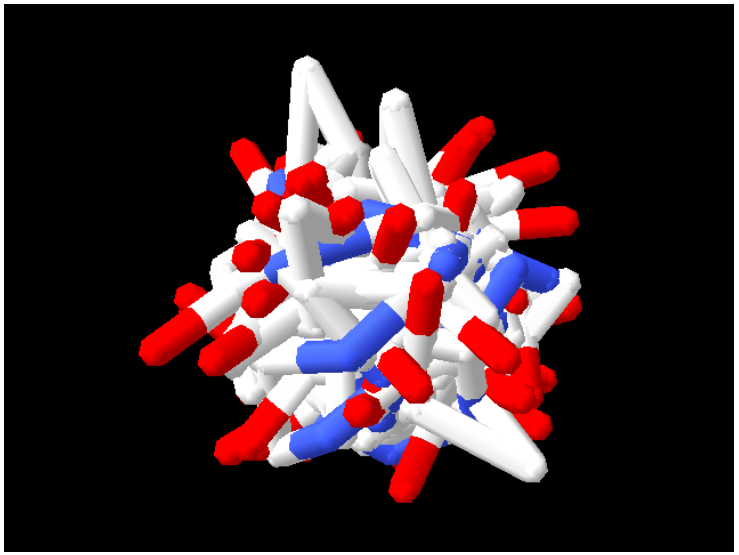
**Experiment 1:** We first examine if the Douglas–Rachford method able to solve the problem, and then investigate the proportion of distances required for a successful reconstruction.

- The protein 1PTQ, whose structure is known, was used.
- Attempt reconstruction using the Douglas–Rachford method from a partial EDM containing the smallest $p$ percent of inter-atomic distances for $p = 1, 2, \ldots, 15$.
- 1,000 iterations performed starting from a random initialization (approx. 2min computation time per instance).

# Experiment 1: Does the Douglas–Rachford Method Work?



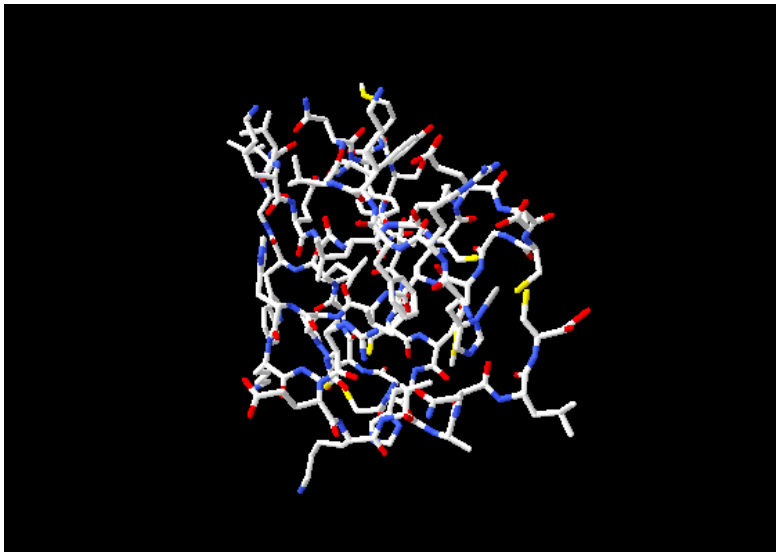
Actual conformation



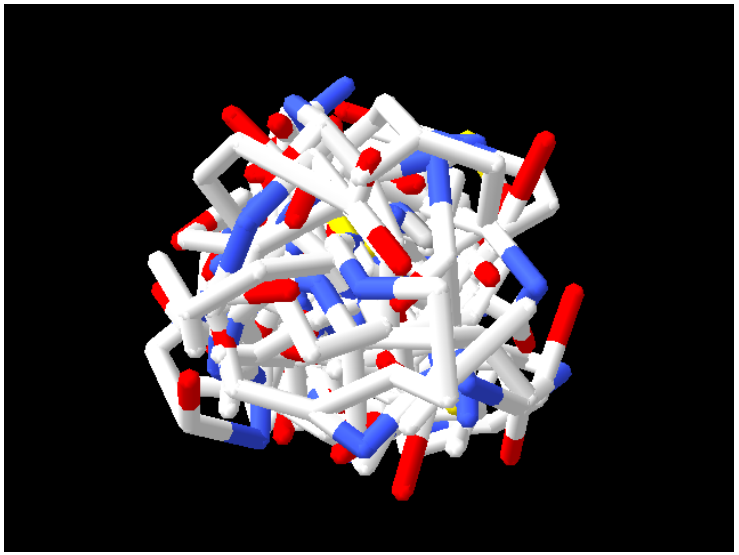
Reconstructed conformation

Distances in partial EDM $= 1\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



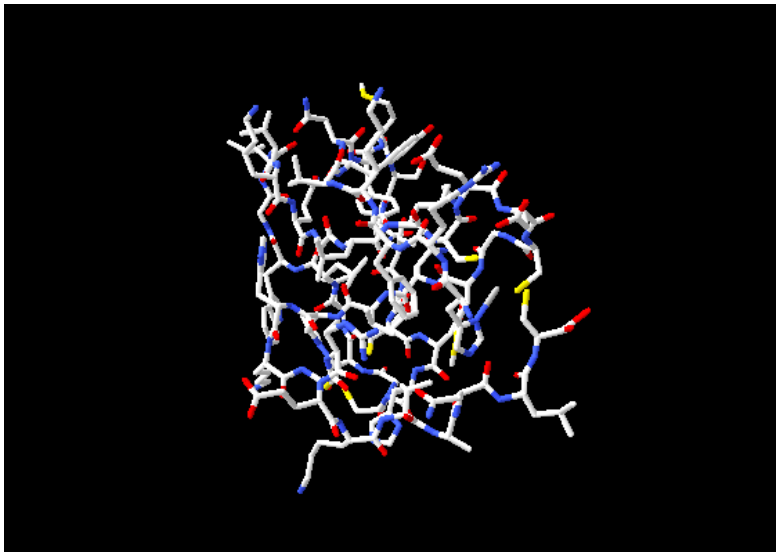
Actual conformation



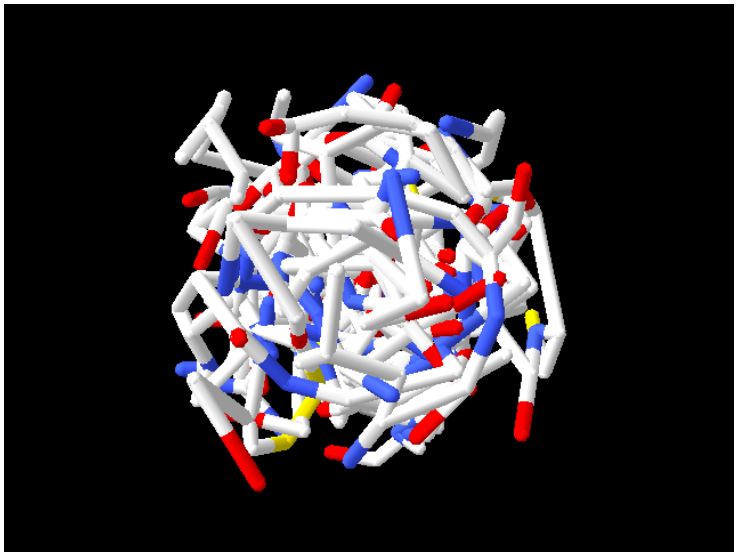
Reconstructed conformation

Distances in partial EDM $= 2\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



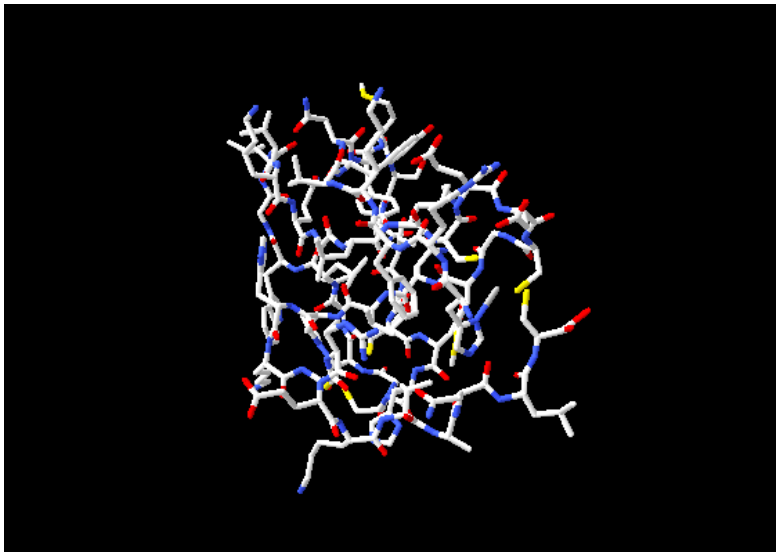
Actual conformation



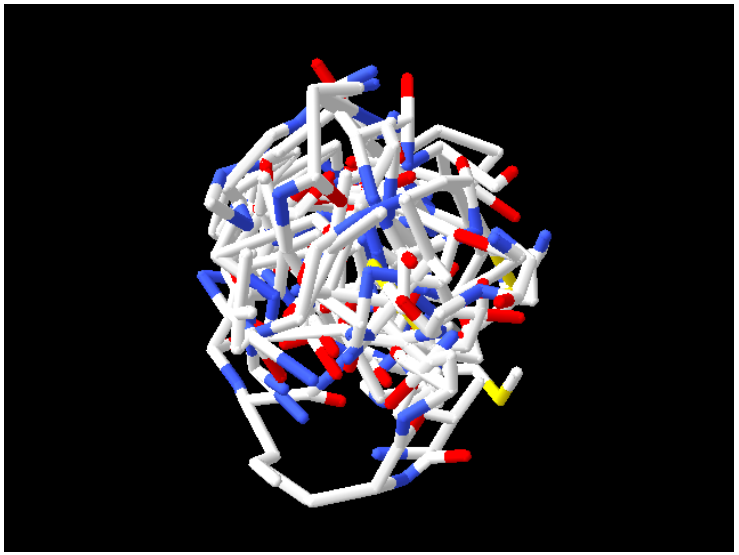
Reconstructed conformation

Distances in partial EDM $= 3\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



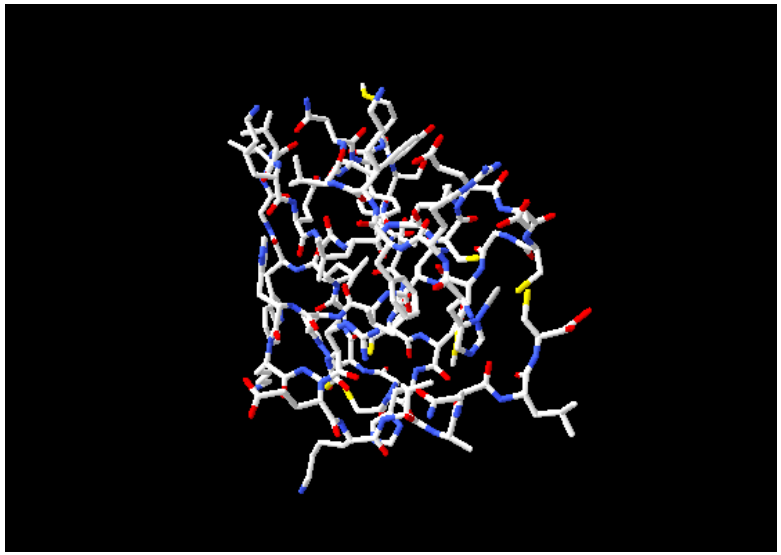
Actual conformation



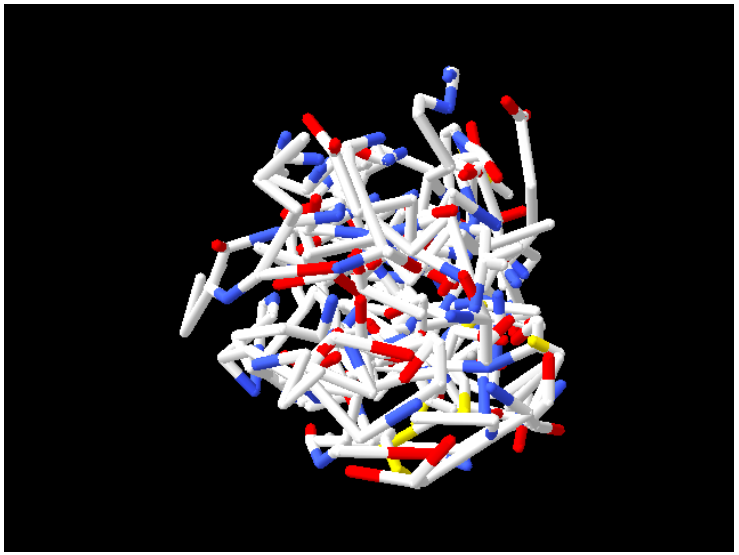
Reconstructed conformation

Distances in partial EDM $= 4\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



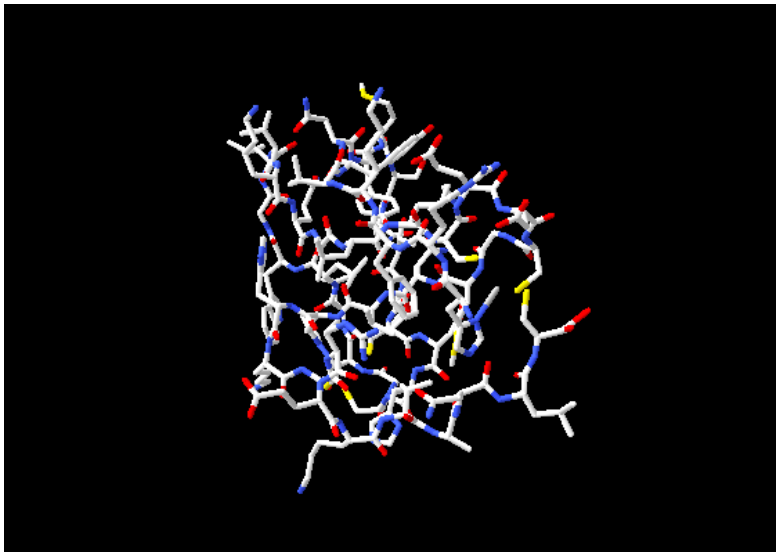
Actual conformation



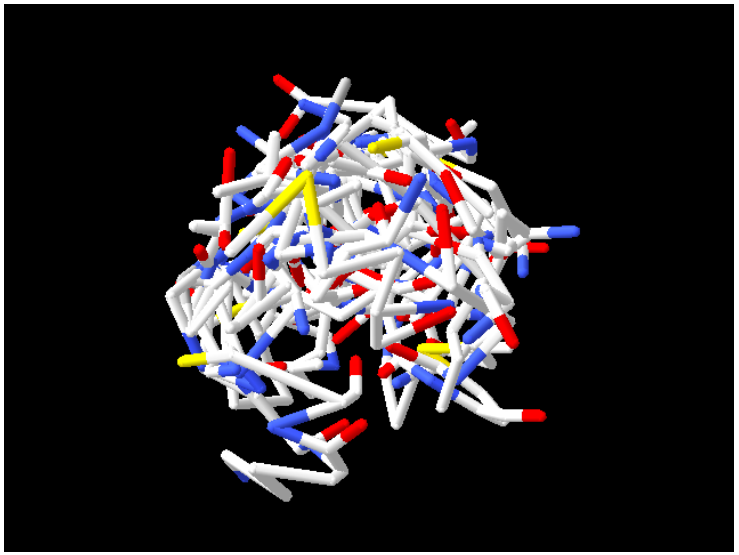
Reconstructed conformation

Distances in partial EDM $= 5\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



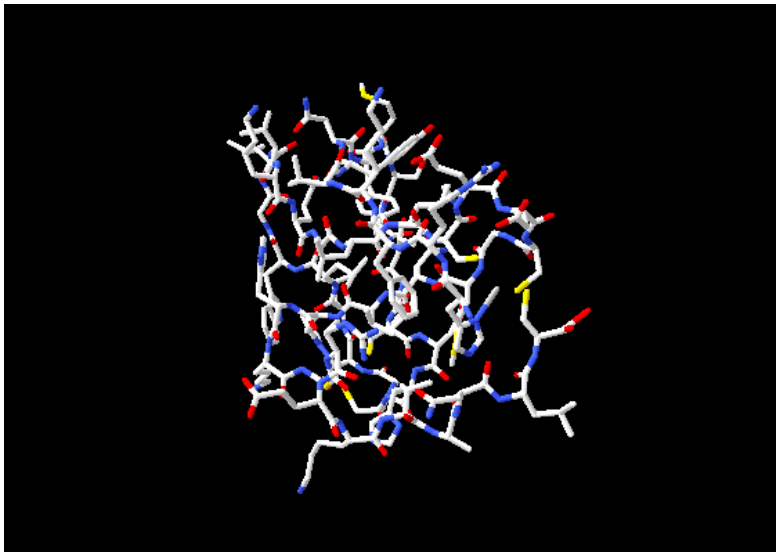
Actual conformation



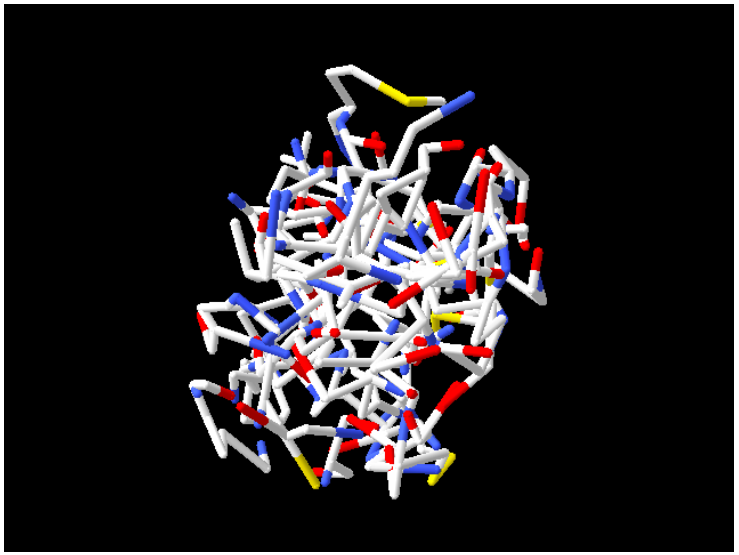
Reconstructed conformation

Distances in partial EDM $= 6\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



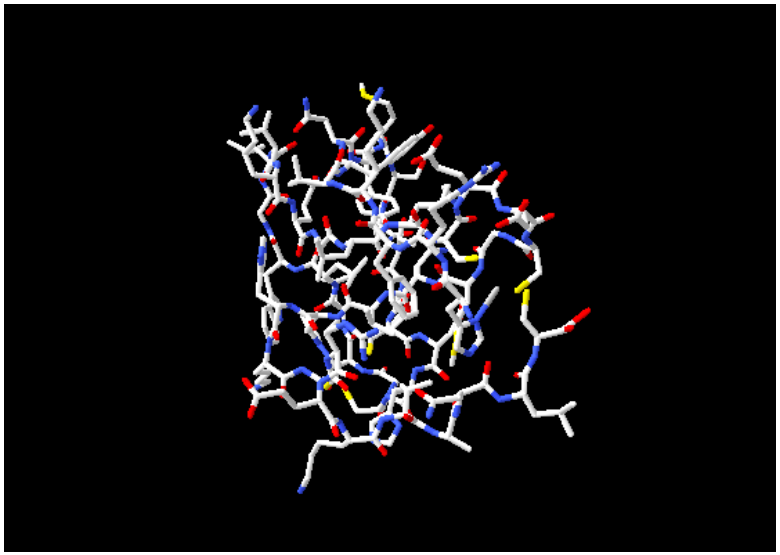
Actual conformation



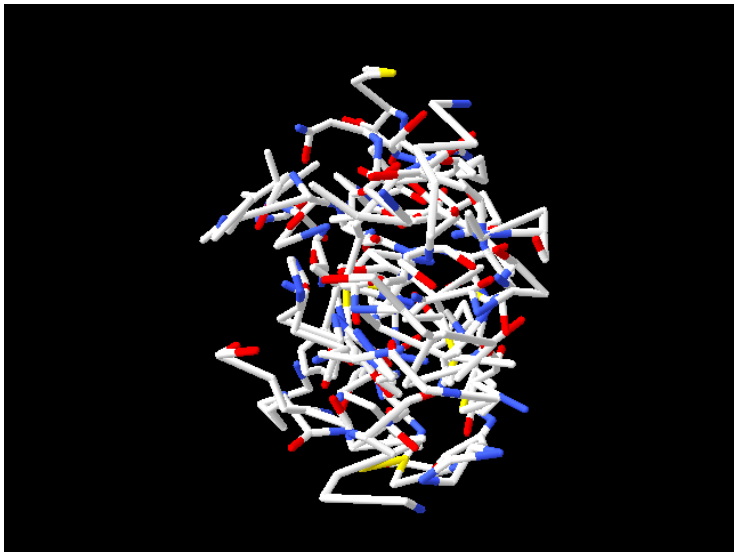
Reconstructed conformation

Distances in partial EDM $= 7\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



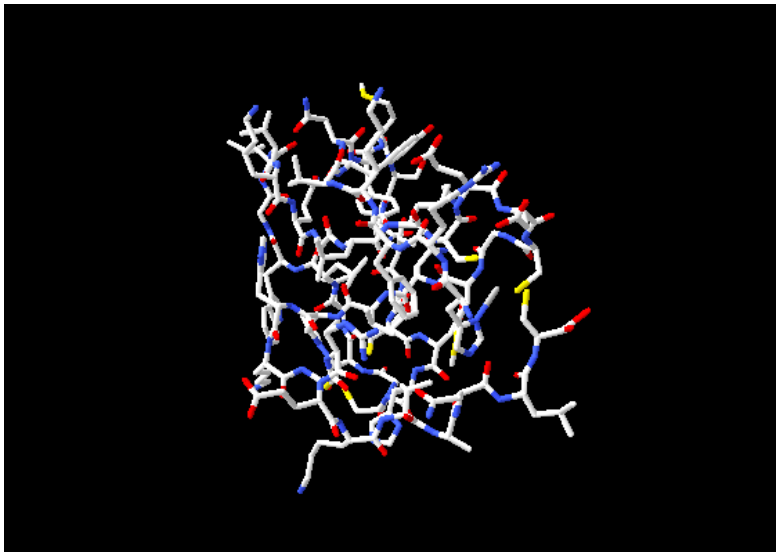
Actual conformation



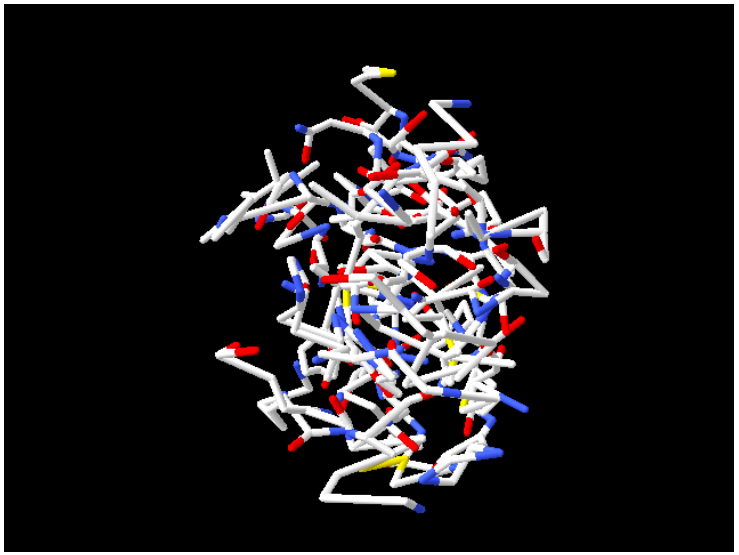
Reconstructed conformation

Distances in partial EDM $= 8\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



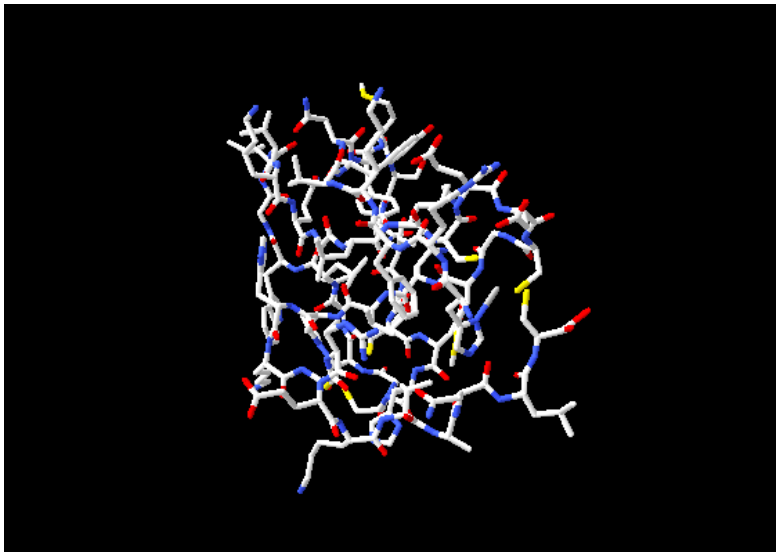
Actual conformation



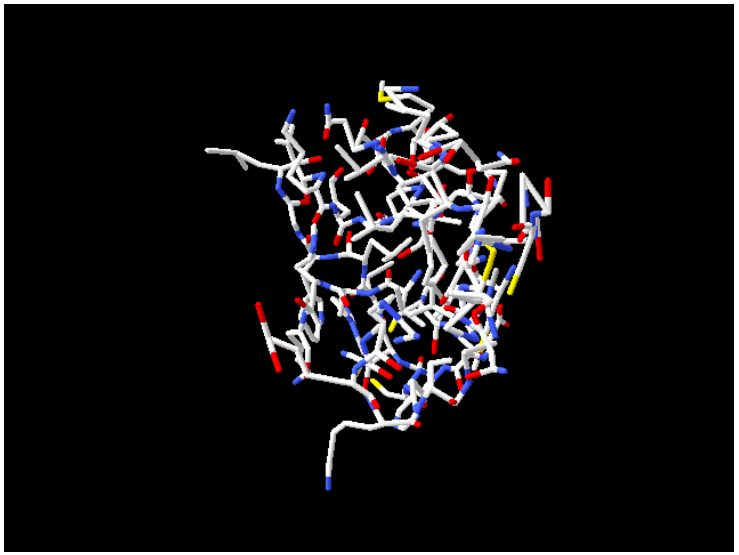
Reconstructed conformation

Distances in partial EDM $= 9\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



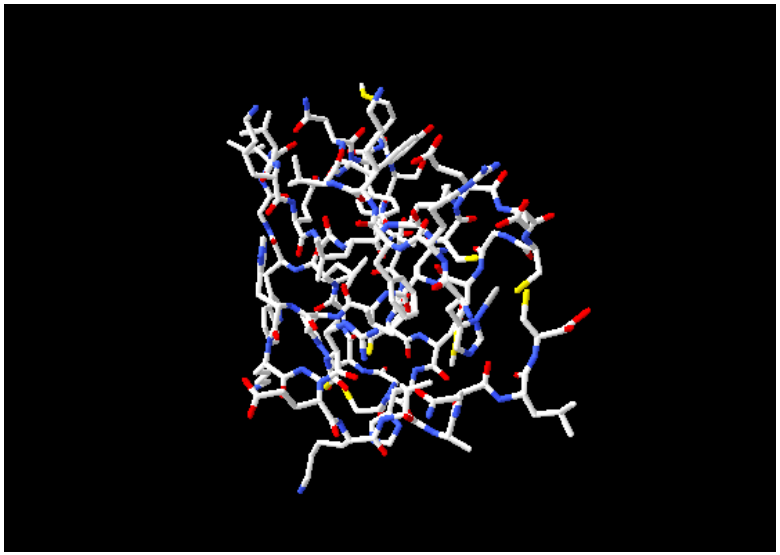
Actual conformation



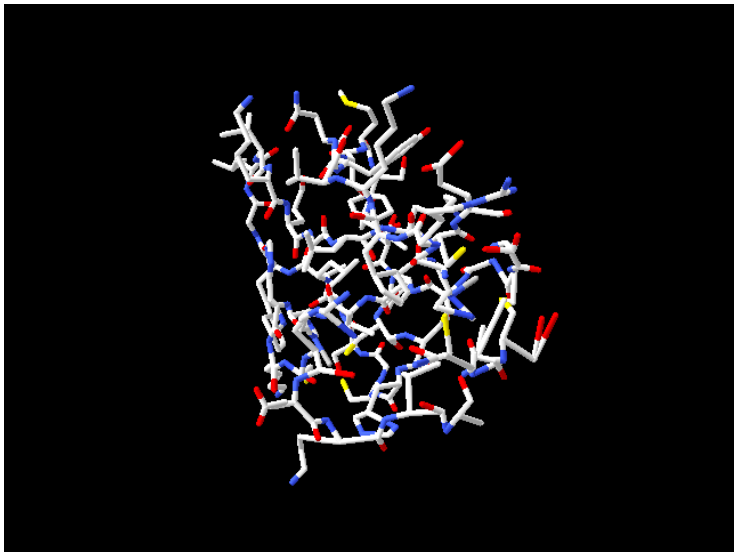
Reconstructed conformation

Distances in partial EDM $= 10\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



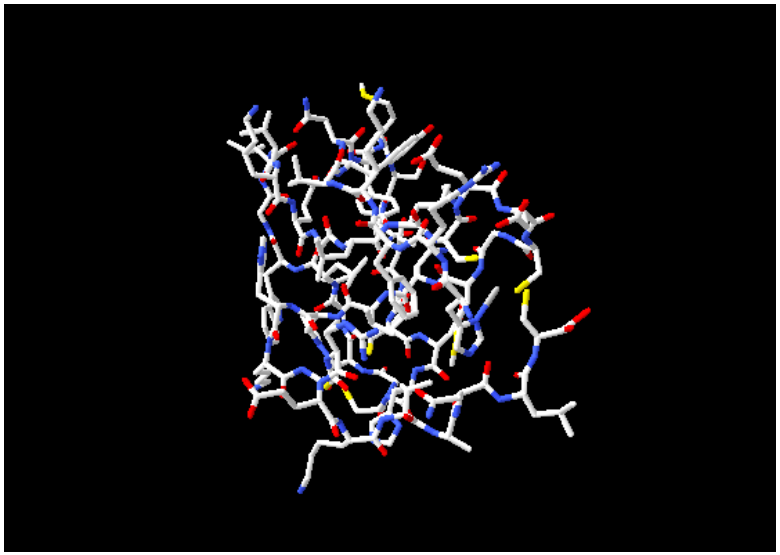
Actual conformation



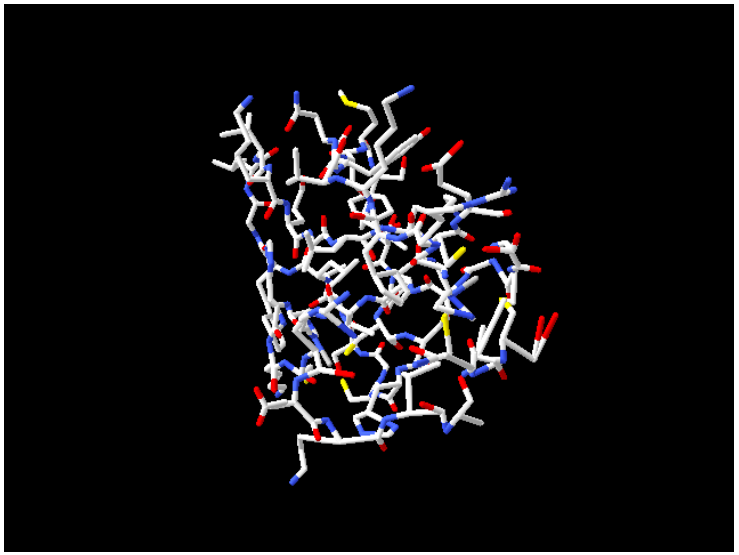
Reconstructed conformation

Distances in partial EDM $= 11\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



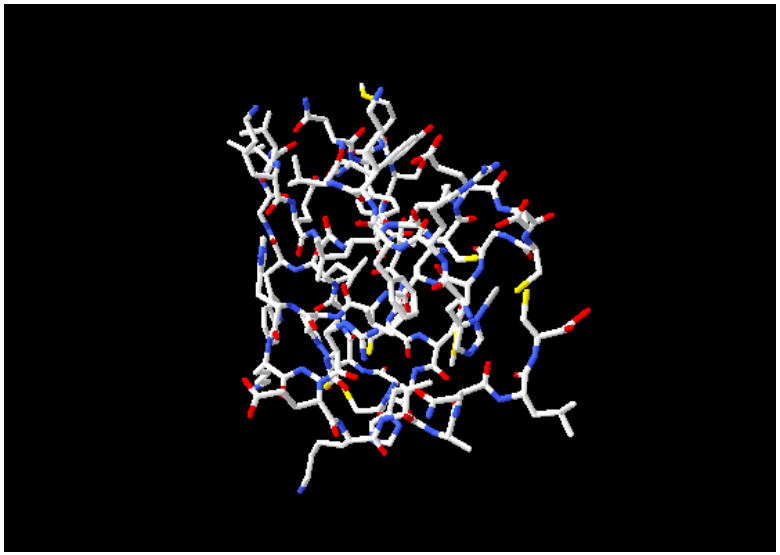
Actual conformation



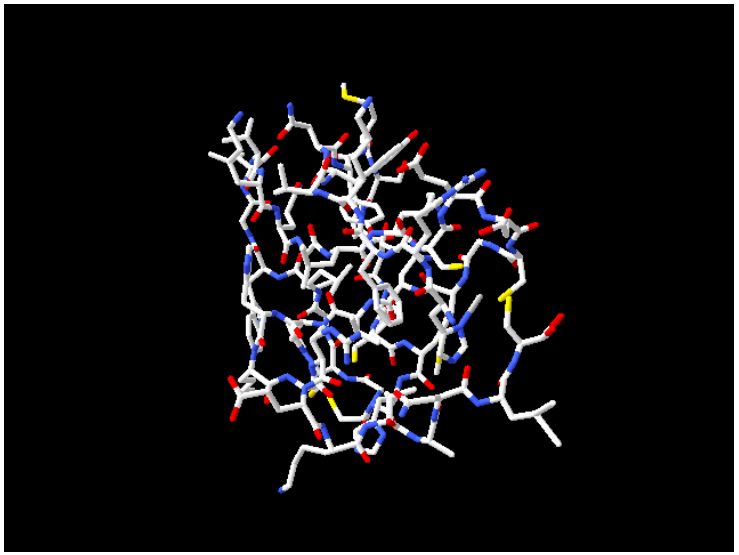
Reconstructed conformation

Distances in partial EDM $= 12\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



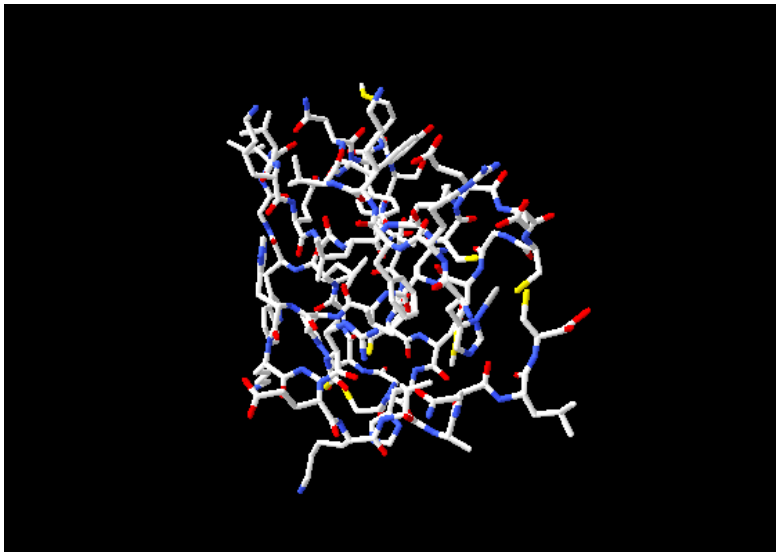
Actual conformation



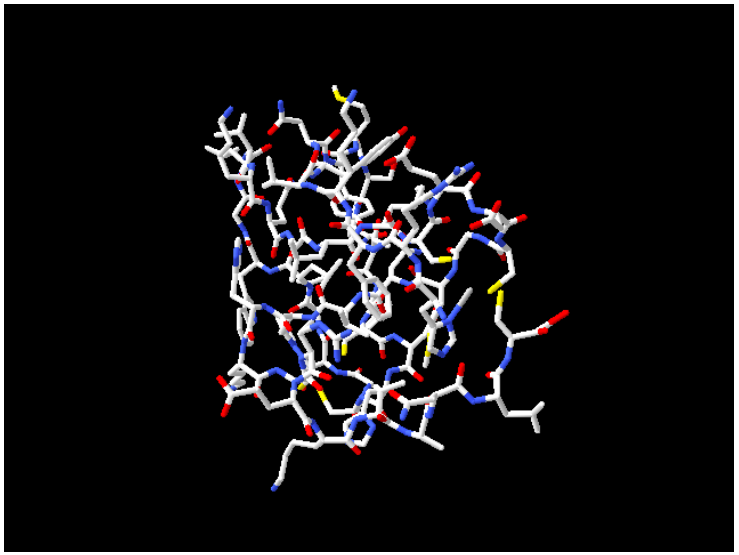
Reconstructed conformation

Distances in partial EDM $= 13\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



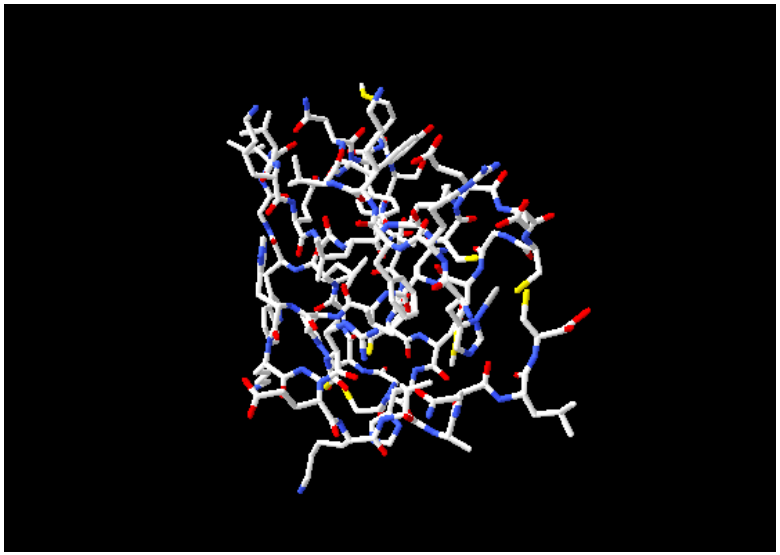
Actual conformation



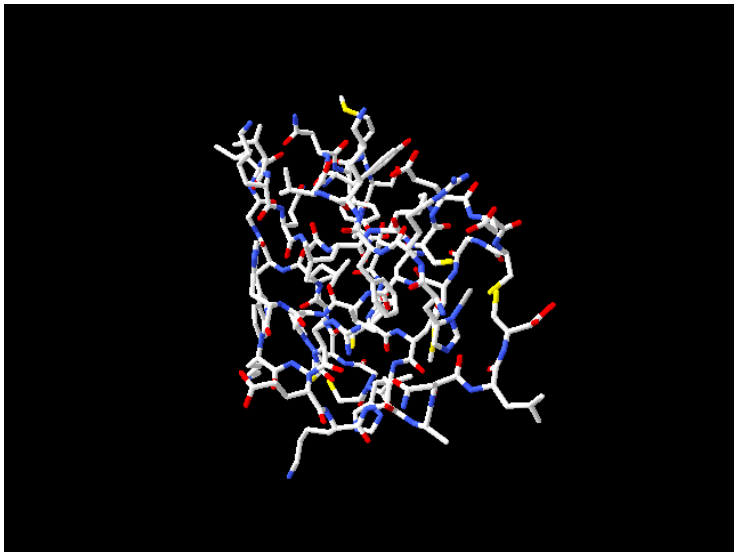
Reconstructed conformation

Distances in partial EDM $= 14\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?



Actual conformation



Reconstructed conformation

Distances in partial EDM $= 15\%$.

# Experiment 1: Does the Douglas–Rachford Method Work?

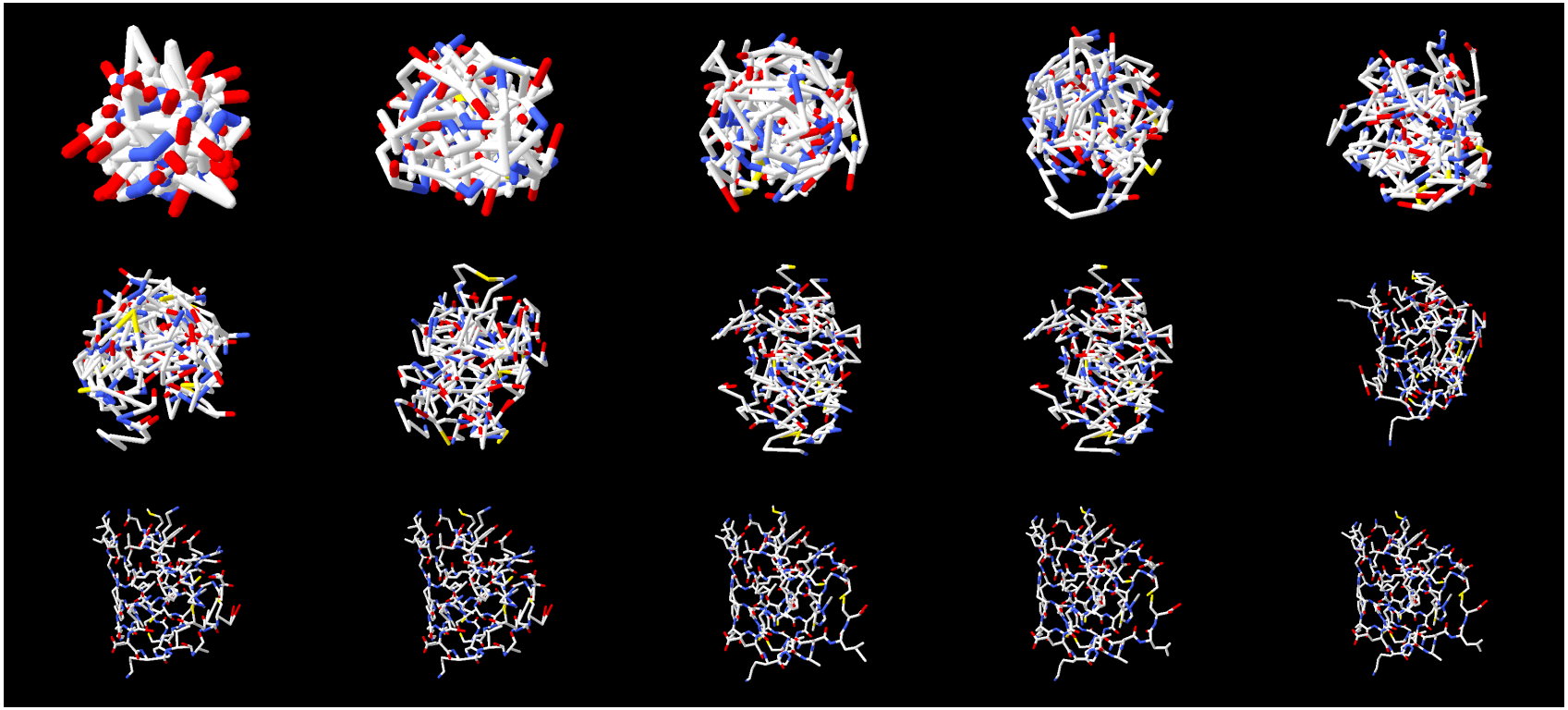


Figure : The reconstructions of 1PTQ. The top-left conformation was obtained from 1% of distances, and the bottom-right from 15% of distances.

# Experiment 1: Does the Douglas–Rachford Method Work?

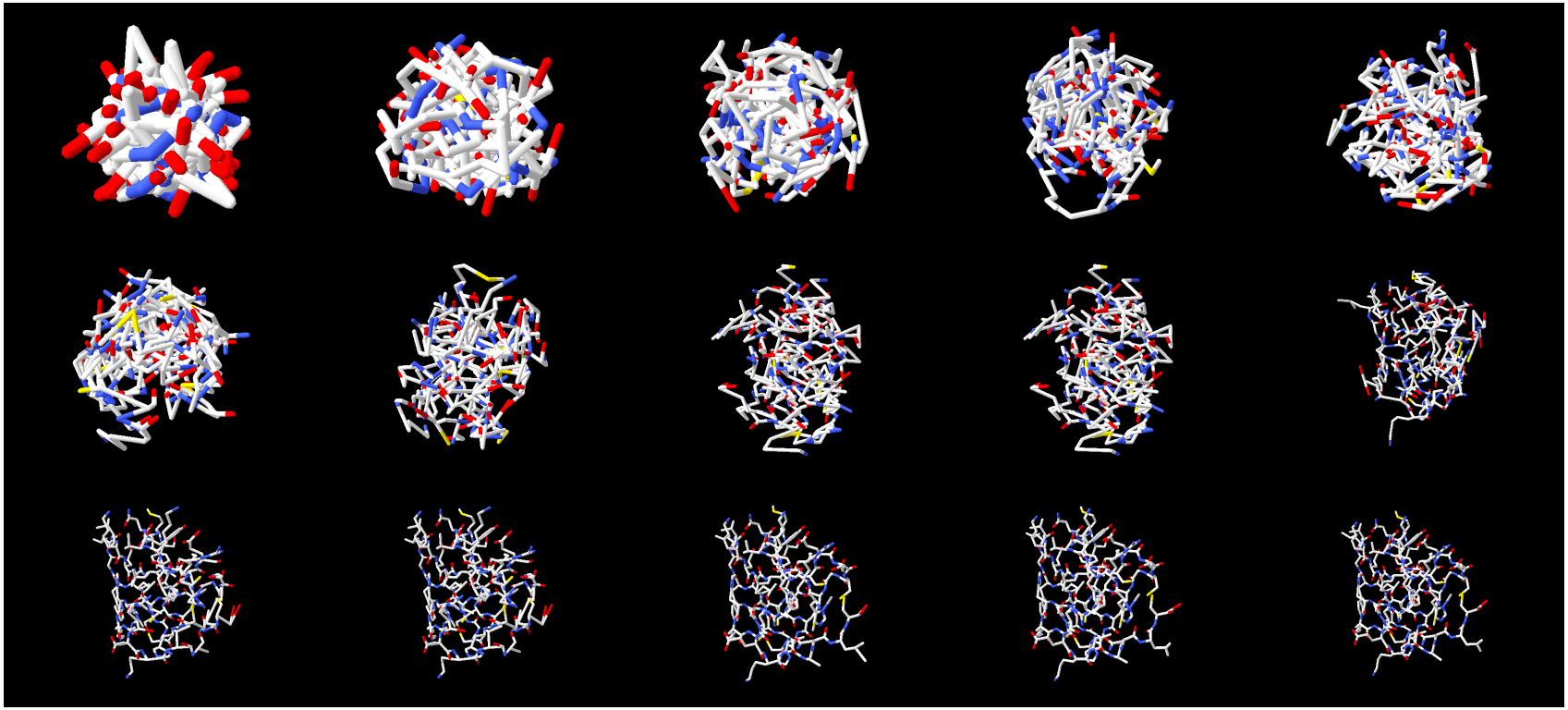


Figure : The reconstructions of 1PTQ. The top-left conformation was obtained from 1% of distances, and the bottom-right from 15% of distances.

- Reconstruction seem possible. For 1,000 iterations approx. 10% of the total (non-zero) distances are needed.

# Experiment 2: Six Test Proteins

**Experiment 2:** We consider the simplest realistic protein conformation determination problem.

NMR experiments were simulated for proteins with known conformation by computing the partial EDM containing all inter-atomic distances $< 6$Å.

Table : Six proteins from the RCSB Protein Data Bank.[2]

| Protein | # Atoms | # Residues | Known Distances |
|---|---|---|---|
| 1PTQ | 404 | 50 | 8.83% |
| 1HOE | 581 | 74 | 6.35% |
| 1LFB | 641 | 99 | 5.57% |
| 1PHT | 988 | 85 | 4.57% |
| 1POA | 1067 | 118 | 3.61% |
| 1AX8 | 1074 | 146 | 3.54% |

[2] `http://www.rcsb.org/`

# Experiment 2: Six Test Proteins

Table : Average (worst) results: 5,000 iterations, five random initializations.

| Protein | Problem Size | Rel. Error (dB) | RMS Error | Max Error |
|---|---|---|---|---|
| 1PTQ | 81,406 | -83.6 (-83.7) | 0.02 (0.02) | 0.08 (0.09) |
| 1HOE | 168,490 | -72.7 (-69.3) | 0.19 (0.26) | 2.88 (5.49) |
| 1LFB | 205,120 | -47.6 (-45.3) | 3.24 (3.53) | 21.68 (24.00) |
| 1PHT | 236,328 | -60.5 (-58.1) | 1.03 (1.18) | 12.71 (13.89) |
| 1POA | 568,711 | -49.3 (-48.1) | 34.09 (34.32) | 81.88 (87.60) |
| 1AX8 | 576,201 | -46.7 (-43.5) | 9.69 (10.36) | 58.55 (62.65) |

- The reconstructed EDM is compared to the actual EDM using:

$$\text{Relative error (decibels)} = 10\log_{10}\left(\frac{\|P_A x_n - P_B R_A x_n\|^2}{\|P_A x_n\|^2}\right).$$

- The reconstructed points in $\mathbb{R}^3$ are then compared using:

$$\text{RMS Error} = \left(\sum_{k=1}^{m}\|z_k - z_k^{\text{actual}}\|^2\right)^{1/2}, \quad \text{Max Error} = \max_{k=1,\ldots,m}\|z_k - z_k^{\text{actual}}\|,$$

which are computed up to translation, reflection and rotation.

# Experiment 2: Six Test Proteins

Table : Average (worst) results: 5,000 iterations, five random initializations.

| Protein | Problem Size | Rel. Error (dB) | RMS Error | Max Error |
|---|---|---|---|---|
| 1PTQ | 81,406 | -83.6 (-83.7) | 0.02 (0.02) | 0.08 (0.09) |
| 1HOE | 168,490 | -72.7 (-69.3) | 0.19 (0.26) | 2.88 (5.49) |
| 1LFB | 205,120 | -47.6 (-45.3) | 3.24 (3.53) | 21.68 (24.00) |
| 1PHT | 236,328 | -60.5 (-58.1) | 1.03 (1.18) | 12.71 (13.89) |
| 1POA | 568,711 | -49.3 (-48.1) | 34.09 (34.32) | 81.88 (87.60) |
| 1AX8 | 576,201 | -46.7 (-43.5) | 9.69 (10.36) | 58.55 (62.65) |

- The reconstructed EDM is compared to the actual EDM using:

$$\text{Relative error (decibels)} = 10 \log_{10} \left( \frac{\|P_A x_n - P_B R_A x_n\|^2}{\|P_A x_n\|^2} \right).$$

- The reconstructed points in $\mathbb{R}^3$ are then compared using:

$$\text{RMS Error} = \left( \sum_{k=1}^{m} \|z_k - z_k^{\text{actual}}\|^2 \right)^{1/2}, \quad \text{Max Error} = \max_{k=1,\ldots,m} \|z_k - z_k^{\text{actual}}\|,$$

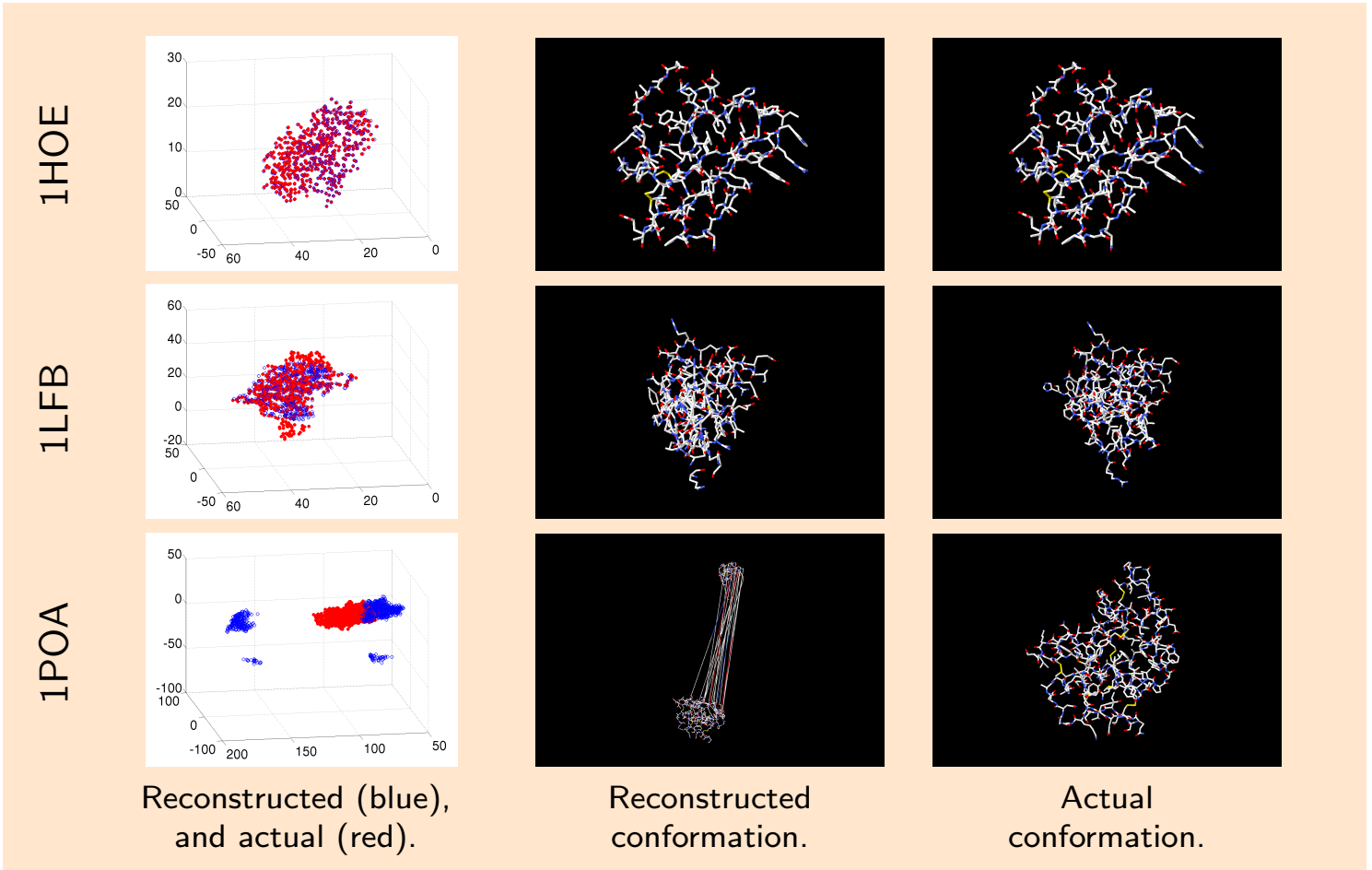
which are computed up to translation, reflection and rotation.

# Experiment 2: Six Test Proteins

How do the errors from the previous table compare to our expectations?

# Experiment 2: Six Test Proteins

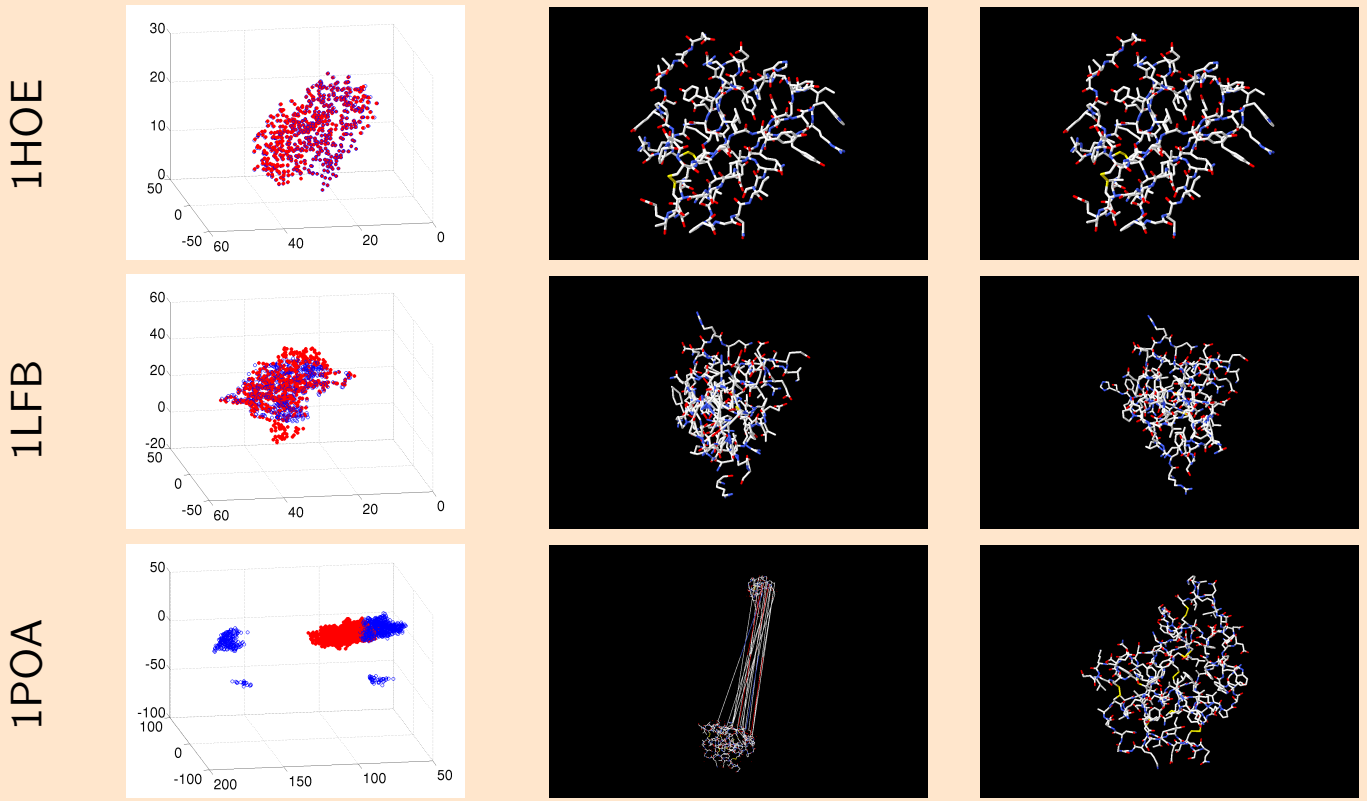
How do the errors from the previous table compare to our expectations?




1HOE is good, 1LFB is mostly good, and 1POA has two good pieces.

# Experiment 2: Six Test Proteins

How do the errors from the previous table compare to our expectations?



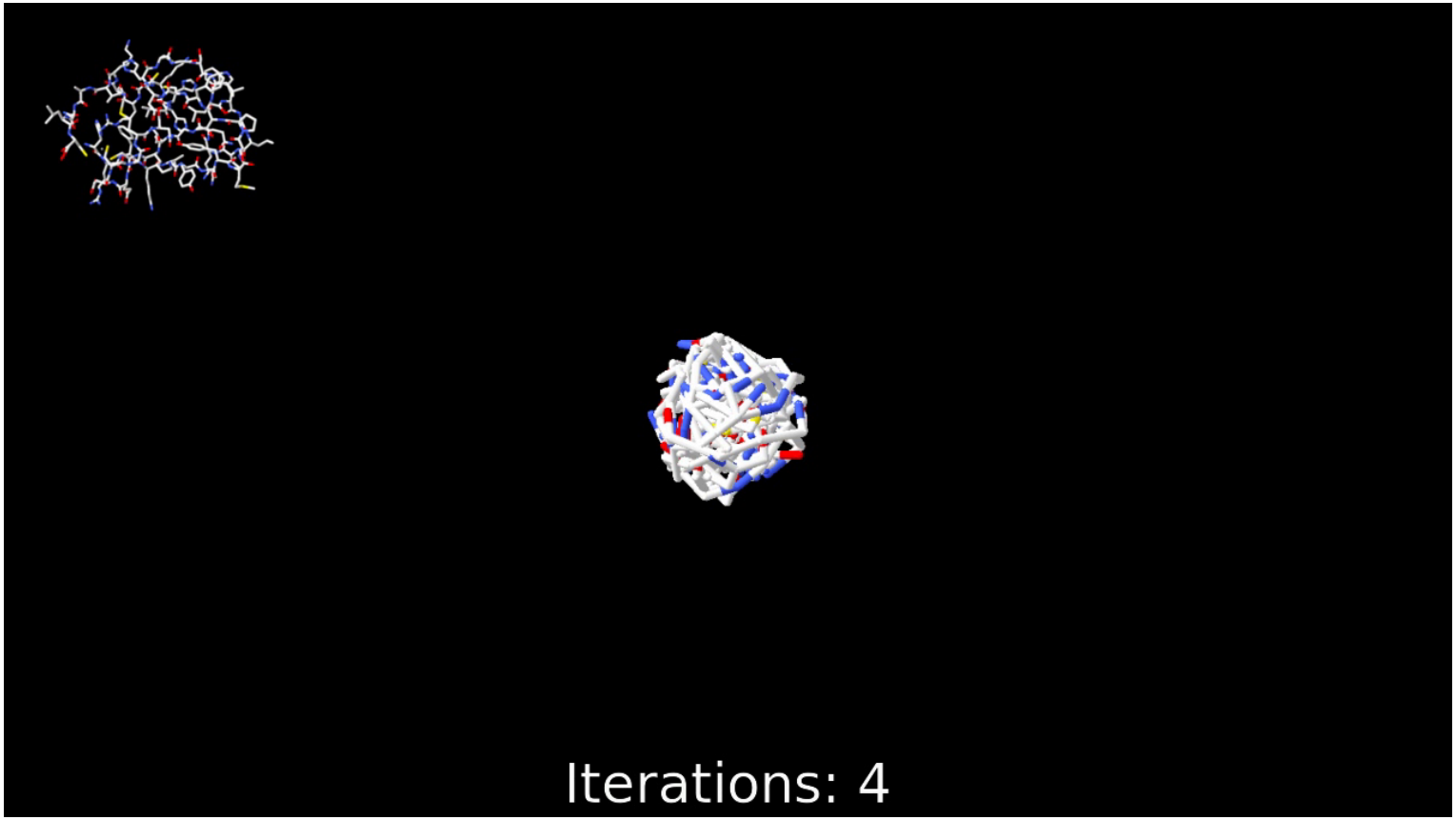

1HOE is good, 1LFB is mostly good, and 1POA has two good pieces.

- Error metrics don't tell the whole story.

# Experiment 3: Why Use the Douglas–Rachford Method?

**Experiment 3:** There are many projection methods, so why should we use the Douglas–Rachford method?

# Experiment 3: Why Use the Douglas–Rachford Method?




First 3,000 steps of the 1PTQ reconstruction

# Experiment 3: Why Use the Douglas–Rachford Method?

**Experiment 3:** There are many projection methods, so why should we use the Douglas–Rachford method?

# Experiment 3: Why Use the Douglas–Rachford Method?

**Experiment 3:** There are many projection methods, so why should we use the Douglas–Rachford method?

A simpler projection method is the method of alternating projections. Given a point $y_0 \in \mathcal{H}$ is given by the fixed-point iteration

$$y_{n+1} := P_{C_2} P_{C_1} y_n.$$

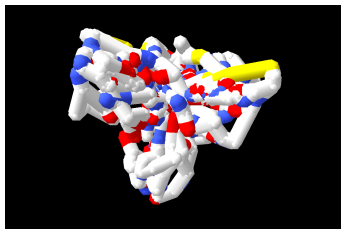
# Experiment 3: Why Use the Douglas–Rachford Method?

**Experiment 3:** There are many projection methods, so why should we use the Douglas–Rachford method?

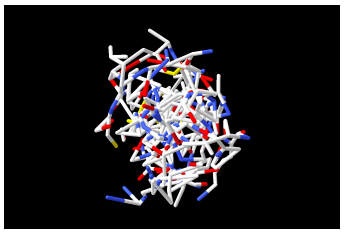
A simpler projection method is the method of alternating projections. Given a point $y_0 \in \mathcal{H}$ is given by the fixed-point iteration

$$y_{n+1} := P_{C_2} P_{C_1} y_n.$$

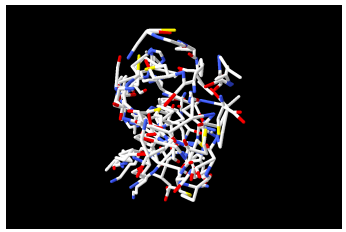
Before reconstruction



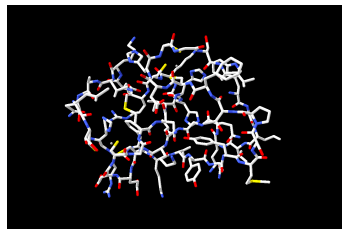
Douglas–Rachford method reconstruction:



500 steps, -25 dB

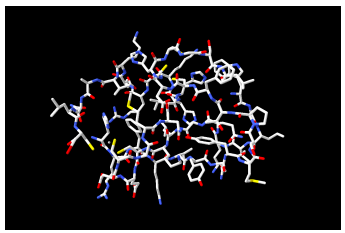


1,000 steps,-30 dB

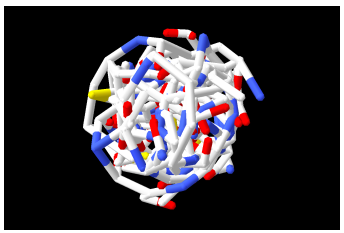


2,000 steps, -51 dB

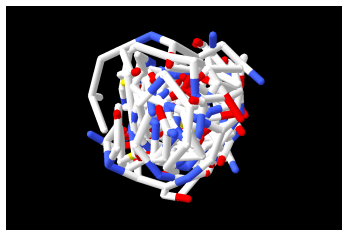
Actual Structure



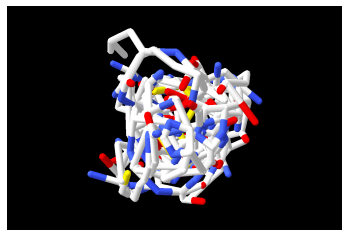
Method of alternating projections reconstruction:



500 steps,-22 dB



1,000 steps, -24 dB



2,000 steps, -25 dB

# Experiment 3: Why Use the Douglas–Rachford Method?

Recall from before:

**Theorem (Basic behaviour of the Douglas–Rachford method)**

Suppose $C_1, C_2$ are closed convex subsets of a finite dimensional Hilbert space $\mathcal{H}$. For any $x_0 \in \mathcal{H}$, define $x_{n+1} = T_{C_1,C_2} x_n$.

1. If $C_1 \cap C_2 \neq \emptyset$, then $x_n \to x$ such that $P_{C_1} x \in C_1 \cap C_2$.
2. If $C_1 \cap C_2 = \emptyset$, then $\|x_n\| \to +\infty$.

The corresponding theorem for alternating projections is:

**Theorem (Basic behaviour of the method of alternating projections)**

Suppose $C_1, C_2$ are closed convex subsets of a finite dimensional Hilbert space $\mathcal{H}$. For any $y_0 \in \mathcal{H}$, define $y_{n+1} = P_{C_2} P_{C_1} y_n$.

1. If $C_1 \cap C_2 \neq \emptyset$, then $y_n \to y \in C_1 \cap C_2$.
2. If $C_1 \cap C_2 = \emptyset$, then $\|P_{C_1} y_n - y_n\| \to d(C_1, C_2)$.

# Concluding Remarks and Future Work

- The Douglas–Rachford method can predict protein conformation using only short-range distances. It performs better than theory suggests.
- Local convergence results for this problem seems possible.
- Alternatively, can the method's behaviour be explained by a CAT(0) metric space interpretation?
- The Douglas–Rachford method is a general purpose algorithm. Can problem specific improvements of the method which exploit special structure present in our constraint sets be made?
- What other applications are fruitful? We are currently investigating an analogous problem of bulk structure determination arising in the context of ionic liquid chemistry.

---

**Douglas–Rachford feasibility methods for matrix completion problems** with F.J. Aragón Artacho & J.M. Borwein. *ANZIAM J.*, accepted 2014. arXiv:1312.7323

Many resources can be found at the companion website:

http://carma.newcastle.edu.au/DRmethods/

# Concluding Remarks and Future Work

- The Douglas–Rachford method can predict protein conformation using only short-range distances. It performs better than theory suggests.
- Local convergence results for this problem seems possible.
- Alternatively, can the method's behaviour be explained by a CAT(0) metric space interpretation?
- The Douglas–Rachford method is a general purpose algorithm. Can problem specific improvements of the method which exploit special structure present in our constraint sets be made?
- What other applications are fruitful? We are currently investigating an analogous problem of bulk structure determination arising in the context of ionic liquid chemistry.

*When presented with a feasibility problem, it is well worth seeing if the Douglas–Rachford method can deal with it – the method is conceptually simple and easy to implement.*

---

**Douglas–Rachford feasibility methods for matrix completion problems** with F.J. Aragón Artacho & J.M. Borwein. *ANZIAM J.*, accepted 2014. arXiv:1312.7323

Many resources can be found at the companion website:

http://carma.newcastle.edu.au/DRmethods/