# ALPHA HELIX PREDICTION BASED ON METROPOLIS-HASTINGS SAMPLING

*Lingyu Ma*[⋆]*, Marco Reisert*[†]*, Christophe Wirth*[‡]*, Carola Hunte*[‡*]*, Olaf Ronneberger*[⋆⋆]*, Hans Burkhardt*[⋆⋆]

[⋆] Department of Computer Science, University of Freiburg, Germany
[†] Department of Radiology, Medical Physics, University Hospital Freiburg, Germany
[‡] Institute for Biochemistry and Molecular Biology, ZBMZ, University of Freiburg, Germany
[*] BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany

## ABSTRACT

Accurate identification of protein secondary structures is beneficial to the structural interpretation of low-resolution X-ray and EM electron density maps. Existing alpha helix identification methods mainly focus on locally voxelwise classification and then link the helix-voxels based on post-processing processes. In this paper, a novel alpha helix identification approach, named as *SSEPredictor*, based on Metropolis-Hastings sampling is proposed, which can provide both locally and globally optimized prediction for alpha helical structures in low-resolution electron density maps. The approach has been tested on X-ray crystallographic electron density maps at 8Å resolution. The experimental results show that the identification accuracy is promising.

***Index Terms***— secondary structure identification, low-resolution X-ray crystallographic data, Metropolis-Hastings sampling, steerable filters.

## 1. INTRODUCTION

The knowledge of the three-dimensional structure of a biological macromolecule is a key step for understanding its function. X-ray crystallography [1] and nuclear magnetic resonance (NMR) [2] are two widely used techniques to obtain high (up to atomic) resolution structures of proteins. Electron microscopy (EM) serves as a complementary tool to study large complexes and macromolecules that are difficult to crystallize and beyond the size threshold for NMR spectroscopy. Well-ordered crystals are a pre-requisite for high-resolution X-ray structures. They are difficult to obtain for membrane proteins and for large, multi-subunit complexes, yet, initial crystals with less order can often be grown. X-ray diffraction data and phases of those crystals can be used for structural analysis at low resolution. Recently, the number of low-resolution X-ray crystallography data deposited in the Protein Data Bank (PDB) [3] has increased [4], strengthening the view that valuable biological information can already be obtained from structural analysis at low resolution.

At intermediate resolutions of 5-10Å, the secondary structure elements (SSEs), $\alpha$-helices and $\beta$-sheets, are visible in the electron density maps of macromolecules. Until recently, several approaches have been developed for EM analysis to identify alpha helices in medium resolution electron density maps. However, they can theoretically also be applied to low-resolution X-ray crystallography maps [4]. 3×3 second moments tensor after five-dimensional template-based cross-correlation is used by the *helixhunter* method [5] to locate candidate helical regions. *EMatch* [6] is another method depending on template matching. Local cylinder-like regions are built with a graph-based method, then a linkage stage is applied for linking fractions of the same helix. *HelixTracer* [7] establishes a forest of graph trees based on Sobel convolution kernels, different trees are merged and finally each left tree corresponds to a helix. *SSEHunter* [8] uses pseudoatoms to reduce the data complexity, each pseudoatom is assigned a composite score based on several criterions, finally all the pseudoatoms with high composite scores are manually linked together to represent the helices. The tensor based method [9] selects local critical points with cylinder-like eigenvalues as the seed points. From each seed point, the method traces helix structure in two opposite directions based on 3×3 local structure tensor.

All existing alpha helix identification approaches mainly focus on locally voxelwise classification with various criterions. The identified helix-like voxels are then linked or gathered into groups. For each group a helix model, usually a line segment with two end points is generated with certain post-processing. The main disadvantage of this kind of method is that the minor errors in local steps can accumulate and affect the final results. In this paper we propose a new alpha helix identification method based on Metropolis-Hastings sampling, which can give global optimization for alpha helix configuration. Similar methods have been used for road detection [10] and fiber tracking [11].

The rest of this paper is organized as follows. Section 2 defines our proposed alpha helix prediction model based on Metropolis-Hastings sampling. In Section 3 a new connect proposal has been designed for generating connections between segments. Section 4 shows how to calculate the external energy part of our model in a fast way. The experimen-

tal results for low-resolution X-ray electron density maps has been given in Section 5. Finally, conclusions and future work are presented in Section 6.

## 2. HELIX STRUCTURE MODELING

Enlightened by [11], we denote an electron density map as $D : \mathbb{R}^3 \to \mathbb{R}$ and the model for global helix structure configuration by $M : \mathbb{R}^3 \to \mathbb{R}$, which denotes the union of all the segments. A single segment (particle) is denoted as $S : \mathbb{R}^3 \to \mathbb{R}$, which determined by a spatial position $\mathbf{x} \in R^3$ and an orientation $\mathbf{n} \in S_2$. We want to maximize the posteriori probability $P(M|D)$ to find the most likely model for given density map.

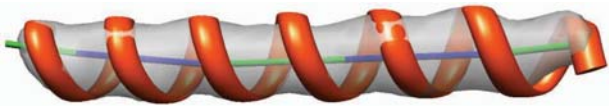
$$M = \sum_{i=1}^{N} S_{\mathbf{x}_i, \mathbf{n}_i}(\mathbf{x}) \tag{1}$$

$$E(M) = \lambda_{\text{in}} E_{\text{in}}(M) + \lambda_{\text{ex}} E_{\text{ex}}(M, D) \tag{2}$$

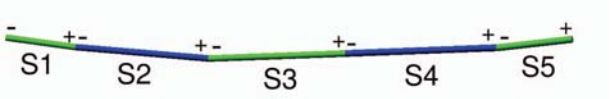
$$P(M|D) = \frac{1}{Z} \exp\left(-\lambda_{\text{in}} \frac{E_{\text{in}}(M)}{T_{\text{in}}} - \lambda_{\text{ex}} \frac{E_{\text{ex}}(M, D)}{T_{\text{ex}}}\right)$$

$$Z = \sum_{M} \exp\left(-\lambda_{\text{in}} \frac{E_{\text{in}}(M)}{T_{\text{in}}} - \lambda_{\text{ex}} \frac{E_{\text{ex}}(M, D)}{T_{\text{ex}}}\right)$$

The external energy $E_{\text{ex}}$ express the dissimilarity between the data and the hypothetical model. The internal energy $E_{\text{in}}$ drives the connected segments into helix-like shape, more precisely the configuration of all segments should be an union of smoothed curves. $\lambda_{\text{in}}$ and $\lambda_{\text{ex}}$ denote the weights for $E_{\text{in}}$ and $E_{\text{ex}}$ respectively. The energy of the configuration $M$ follows Equation (2). $T_{\text{ex}}$ and $T_{\text{in}}$ are global time-varying parameters called the *temperatures*.



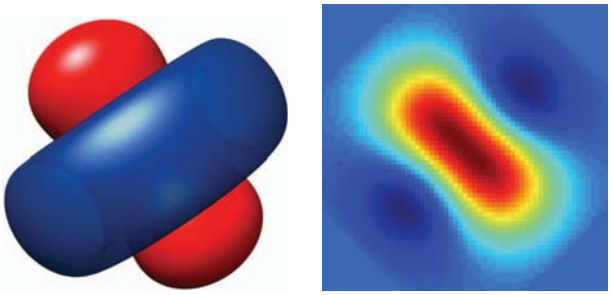
(a) Structure of alpha helix in ribbon presentation and corresponding electron density map




(b) Segments configuration

**Fig. 1**. Expected configuration of a single alpha-helical transmembrane segment of the M2 protein from influenza A virus (PDB entry 1MP6)

The internal energy part of the proposed model for detecting alpha helices in electron density maps is the same as defined in [11]. Figure 1 demonstrates the expected configuration of segments for the transmembrane domain of the M2 protein from influenza A virus (PDB entry 1MP6), which is composed of a single, nearly ideal alpha helix. The electron density map is shown as transparent gray surface in Figure 1(a). There are five segments (S1 to S5) linked together to show the location of the helix axis. The minus and plus signs denote two endpoints of each segment, their positions can be calculated as $\mathbf{x} - l\mathbf{n}$ and $\mathbf{x} + l\mathbf{n}$, where $l$ is half of the length of one single segment. Note all the images in this paper are generated with UCSF Chimera [12].



(a) 3D structure (b) Central slice

**Fig. 2**. Segment demonstration

Figure 2 demonstrates the segment we use for helix prediction. The red and blue parts correspond to positive and negative density respectively. The segment can be generated as

$$S_o = -\frac{2}{3} g_{\text{xx}} + g_{\text{yy}} + g_{\text{zz}}$$

$$S = \frac{-w}{\max(S_o)} S_o \tag{3}$$

where $g_{\text{xx}}$, $g_{\text{yy}}$ and $g_{\text{zz}}$ are secondary derivatives of gaussian function. $\max(S_o)$ denotes the maximum of structure $S_o$, which has been presented as the optimal steerable filter for curve detection [13]. $w$ is the weight of the segment.

To maximize $P(M|D)$, the Metropolis Hastings sampler [14] has been used to draw samples from the posterior distribution. By lowering the temperatures $T_{\text{in}}$ and $T_{\text{ex}}$, it becomes more and more likely to sample from maxima of $P(M|D)$ which correspond to minima of $E(M)$. The idea of the Metropolis Hastings sampler is to choose certain state transitions randomly. These transitions are chosen according to certain proposal distributions. The approach is as follows: choose a modification of the current state $M$ according to a proposal distribution $p^{\text{prop}}$, we call this modification $M'$. Then, accept this modification if the so-called Green's ratio $R$ is above 1, where $R$ is given by

$$R = \frac{P(M' \mid D)}{P(M \mid D)} \frac{p^{\text{prop}}(M \mid M')}{p^{\text{prop}}(M' \mid M)}. \tag{4}$$

If $R$ is below 1, the modification is accepted with probability $R$. After a certain number of iterations (the burn-in

phase) the resulting chain of states follows the desired distribution. Then, the temperature is slowly decreased. The proposal splits into three different types, where each of the proposal is selected with a certain probability: segment creation/deletion ($p_{\text{birth}}, p_{\text{death}}$), segment moves ($p_{\text{shift}}, p_{\text{opt}}$) and segment connections ($p_{\text{conn}}$). Pseudocode of the algorithm is given below.

**Algorithm 1** Optimization

**Input: Start with empty reconstruction** $\mathcal{M}$
**Output: helix reconstruction** $\mathcal{M}$

1: **for** $j = 1 : \#iterations$ **do**
2: Select one of the proposals according to the probabilities $p_{\text{birth}}, p_{\text{death}}, p_{\text{shift}}, p_{\text{opt}}, p_{\text{conn}}$
3: According to proposal type, generate new state $M'$
4: Compute Green's ratio according to equation (4)
5: **if** $R >$ random number $\in [0, 1]$ **then**
6: $M := M'$
7: **end if**
8: Reduce the temperature
9: **end for**

## 3. CONNECT PROPOSAL

Segment creation/deletion and moves mentioned in the previous section follow the techniques proposed in [11]. Here we propose a new connect strategy, which has more concise definition and needs less computation time. Once the connection proposal is chosen, select uniformly an endpoint of a segment. Call this endpoint $e$. Thus, the particular endpoint $e$ is selected with probability $\frac{1}{2N}$ where $N$ is the number of segments in the current state.

If the endpoint is free, then propose a connection that for all neighboring endpoints in the neighborhood of $e$ compute

$$p_e(e_k) = \frac{1}{Z} \exp(-U^p_{\text{con}}(e, e_k)/T_{\text{prop}})$$

where

$$Z = \sum_{k \in \text{neighbors}} \exp(-U^p_{\text{con}}(e, e_k)/T_{\text{prop}})$$

and $U^p_{\text{con}}$ is the proposal connection potential

$$U^p_{\text{con}}(e, e') = (e - \overline{\mathbf{x}})^2 + (e' - \overline{\mathbf{x}})^2 - L_{\text{prop}}$$

$$\overline{x} = \frac{1}{2}(\mathbf{x}_e + \mathbf{x}_{e'})$$

where $\mathbf{x}_e$ and $\mathbf{x}_{e'}$ denote the center positions of the two segments containing endpoints $e$ and $e'$. Based on this select a neighbor $e_j$ according to the distribution $p_e$. The Gibbs ratio for connecting the endpoint $e$ with this neighbor can be written as

$$\begin{aligned} R &= \exp(-\Delta E^{\text{con}}/T_{\text{in}}) \frac{1}{p_e(e_j)} \\ &= \exp[-U_{\text{con}}(e, e_j)/T_{\text{in}} + U^p_{\text{con}}(e, e_j)/T_{\text{prop}}] \cdot Z \end{aligned}$$

In the special case $T_{\text{in}} = T_{\text{prop}}$ and $L = L_{\text{prop}}$ we obviously have $R = Z$. If just $T_{\text{in}} = T_{\text{prop}}$ then

$$R = \exp(L - L_{\text{prop}}) \cdot Z \tag{5}$$

If the endpoint $e$ is already connected with an endpoint $e_j$ break the connection with Gibbs ratio

$$\begin{aligned} R &= \exp(-\Delta E^{break}/T_{\text{in}}) \cdot p_e(e_j) \\ &= \exp[U_{\text{con}}(e, e_j)/T_{\text{in}} - U^p_{\text{con}}(e, e_j)/T_{\text{prop}}]/Z \end{aligned}$$

If $e_j$ is not in the neighborhood of $e$, which means it has moved out of the neighborhood of $e$, then $p_e(e_j) = 0$, and consequently $R = 0$. It means we are not allowed to break the connection in this situation. Again, for the special case $T_{\text{in}} = T_{\text{prop}}$ and $L = L_{\text{prop}}$ we have $R = 1/Z$. And also, if just $T_{\text{in}} = T_{\text{prop}}$ we have

$$R = \exp(-(L - L_{\text{prop}}))/Z \tag{6}$$

## 4. EXTERNAL ENERGY

The external energy part in [11] is specially designed for High-Angular-Resolution-Diffusion data (HARDI), here we propose a new external energy strategy for low-resolution electron density maps. According to Equation (1), a birth or death of segment can be expressed as

$$M' = \sum_{i=1}^{N} S_{\mathbf{x}_i, \mathbf{n}_i}(\mathbf{x}) \pm S_{\mathbf{x}_j, \mathbf{n}_j}(\mathbf{x}) \tag{7}$$

and it only effects the external energy, which can be expressed as (for simplification the suffix $\mathbf{x}_j$ and $\mathbf{n}_j$ have been replaced by a single $j$)

$$\begin{aligned} \triangle E_{\text{ex}} &= \int |M \pm S_j - D|^2 \, d\mathbf{x} - \int |M - D|^2 \, d\mathbf{x} \\ &= \pm 2 \int M \cdot S_j \, d\mathbf{x} \mp 2 \int D \cdot S_j \, d\mathbf{x} + \int S_j^2 \, d\mathbf{x} \end{aligned}$$

where $S_j$ is just a scalar, $M \cdot S_j$ and $D \cdot S_j$ can be precalculated to generate two look-up tables to reduce the online running time. For $D \cdot S_j$, just define

$$F(\mathbf{x}_j, \mathbf{n}_j) = \int D \cdot S_{\mathbf{x}_j, \mathbf{n}_j} \, d\mathbf{x}$$

when $\mathbf{n}_j$ is given, it can be calculated in a convolution way

$$F_{\mathbf{n}_j}(\mathbf{x}_j) = D * S_{\mathbf{n}_j} = \int D(\tau) S_{\mathbf{n}_j}(\mathbf{x}_j - \tau) \, d\tau$$

For $M \cdot S_j$

$$\int M \cdot S_j \, d\mathbf{x} = \sum_{i=1}^{N} \int S_{\mathbf{x}_i,\mathbf{n}_i}(\mathbf{x}) \cdot S_{\mathbf{x}_j,\mathbf{n}_j}(\mathbf{x}) \, d\mathbf{x}$$

let's define

$$G(\mathbf{x}_i, \mathbf{n}_i, \mathbf{x}_j, \mathbf{n}_j) = \int S_{\mathbf{x}_i,\mathbf{n}_i}(\mathbf{x}) \cdot S_{\mathbf{x}_j,\mathbf{n}_j}(\mathbf{x}) \, d\mathbf{x}$$

if we set $S_i$ fixed, and let $S_j$ move and rotate around it, again $G$ can be calculated in a convolution way

$$G_{\mathbf{n}_i,\mathbf{n}_j}(\overline{\mathbf{x}}) = S_{\mathbf{x}_o,\mathbf{n}_i} * S_{\overline{\mathbf{x}},\mathbf{n}_j} = \int S_{\mathbf{x}_o,\mathbf{n}_i}(\tau) \cdot S_{\overline{\mathbf{x}},\mathbf{n}_j}(\overline{\mathbf{x}} - \tau) \, d\tau$$

where $\overline{\mathbf{x}} = \mathbf{x}_j - \mathbf{x}_i$. The change of external energy for shift and opt-shift proposal [11] can be calculated as a death proposal followed by a birth proposal.

## 5. EXPERIMENTS

### 5.1. Experimental Data

Three protein structures with different SSE composition were selected for validation of the experimental approach. The multidrug ABC transporter (2HYD) consists mainly of alpha-helices. The porin NanC (2WJQ) is an outer membrane protein with the typical beta-barrel fold. The response regulator PleD (2V0N) has a mixed SSE composition. Protein structures as well as the corresponding structure factor files for selected model proteins were taken from the PDB [3]. Electron density maps were calculated using the experimental intensities and structure based phases at 8Å resolution using the CCP4 software package [15]. The X-ray structures and their corresponding electron density maps are displayed in Figure 3(a), 5(a) and 6(a), respectively.

### 5.2. Parameters

The proposed approach depends on several parameters. According to Equation 3, there are two parameters for each segment, the weight $w$ and the standard deviation $\sigma$ of the gaussian function. $\sigma$ controls the diameter of the positive part of each segment, it has been set to 3 according to the mean radius 2.5Å for a typical alpha helix. $w$ should be proportional to the mean of local maximia, here it's set to 0.002. If $w$ is too low, we will have superfluous segments for one helix, if it's too high, false negative detections will occur. The segment length $l$ controls the expected curvature of the fibers. Large $l$ imply low curvature and vice verse. Here $l$ is set to 9. The connection bias $L$, whose definition can be found in [11], is set to 1.5. Large value can cause lots of false positive connections. The weights for internal and external energies are set as $\lambda_{\text{in}} = \lambda_{\text{ex}} = 1$.

Other parameters are the iteration parameters. The individual proposal probabilities are set as follows: $p_{\text{birth}} = 0.45$, $p_{\text{death}} = 0.12$, $p_{\text{shift}} = 0.22$, $p_{\text{opt}} = 0.12$ and $p_{\text{con}} = 0.09$. The prediction results do not depend on these probabilities, they just influence the number of iterations to achieve a stable result. The number of iterations depends mainly on the complexity and the size of the protein electron density map. If the density map is supposed to have many alpha helices, the iterations number should be high. In our experiments, the number of iterations are set to 30 millions. The starting external temperature $T_{\text{ex}}$ is set to 2. Too high starting temperature can cause the sampler stay too long time in the 'burn-in' stage, and too low starting temperature will cause unstable prediction results. The decreasing ratio for $T_{\text{ex}}$ is set to 0.95 for 100 thousand iterations. $T_{\text{in}}$ is set to 1 and fixed during the whole sampling process. Especially $T_{\text{prop}}$ and $L_{\text{prop}}$ are set to equal to $T_{\text{in}}$ and $L$, respectively.

### 5.3. Run time

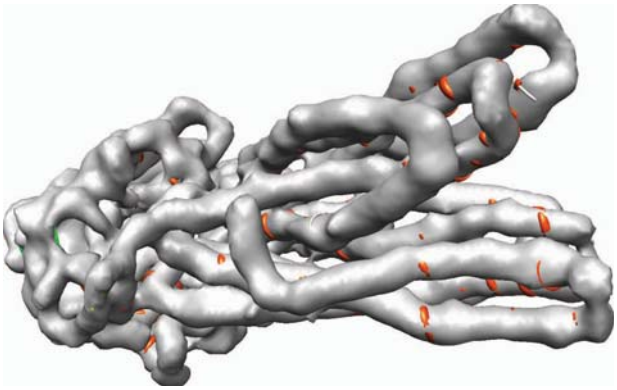
The *SSEPredictor* is implemented on a PC equipped with Intel(R) Core(TM)2 Duo 3.00GHz CPU. The run time of the program is data dependent. Take the multidrug ABC transporter (2HYD) as an example, there are 45059 nontrivial voxels, whose densities are higher than the mean of local maximia, in the electron density map. It takes approximately 8 minutes to generate the final prediction.

### 5.4. Results and discussion

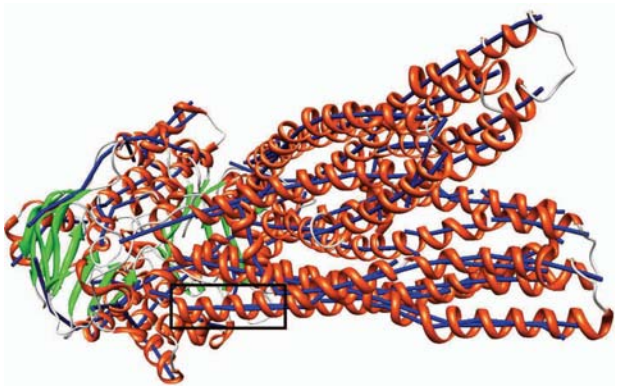
As mentioned in Section 1, there are several methods existing for alpha helix identification for low-resolution electron density maps. In this section the performance of our proposed approach, which we name as *SSEPredictor*, will be compared with two of them, the *HelixTracer* method [7] and the *Localtensor* approach [16]. We do not compare with the *helixhunter* [5], which is integrated in the *SSEHunter* approach [8], because it needs manual interaction. The source code of another automatic method *Ematch* [6], was not available to us.

Figure 3(b-d), 5(b-d) and 6(b-d) show the superimposition of the experimental helix prediction (blue threads) superimposed on the structural model for protein 2V0N, 2WJQ and 2HYD, respectively[1]. For all the three proteins, *SSEPredictor* gives less false positive detections than the other two methods, in particular for the pure beta protein 2WJQ. Sometimes the length of the predicted helix of *SSEPredictor* is longer than the real length of the helix (see the black circle parts of Figure 6(b)). This might be caused by the internal energy, which makes it try to connect as many potential segments as possible. This problem can be solve by adjusting the weight $\lambda_{\text{in}}$ and $\lambda_{\text{ex}}$ for internal energy and external energy. *SSEPredictor* has advantages for detecting long and nonhomogeneous

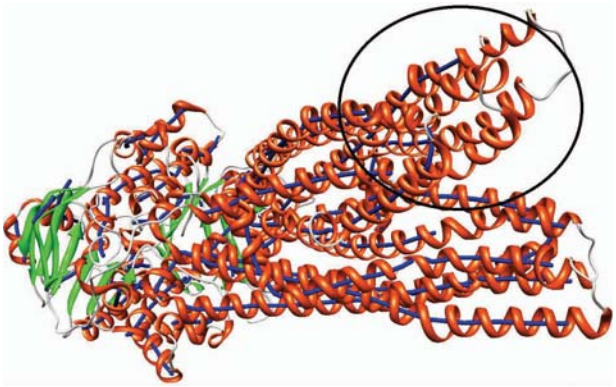
[1]The parameters of *HelixTracer* and *Local tensor* have been optimized according to the prediction results of protein 2V0N and 2HYD by grid search method.



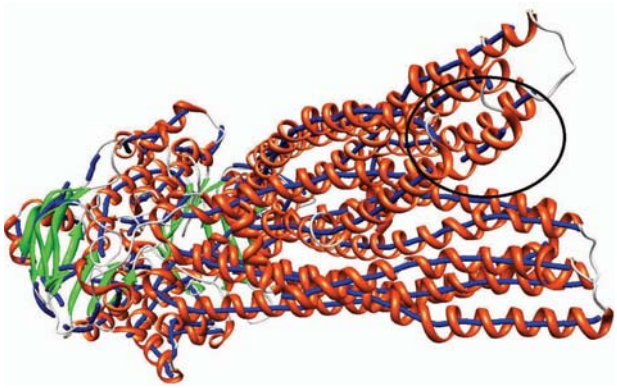
(a) Electron density map



(b) SSEPredictor (our method)



(c) HelixTracer



(d) LocalTensor

**Fig. 3**. The X-ray structures of the multidrug ABC transporter (PDBentry 2HYD) are shown in ribbon presentation with $\alpha$-helices and $\beta$-sheets highlighted in red and green, respectively. The electron density maps are shown in gray and the experimental helix prediction are shown as blue threads.

helices, that is, for *SSEPredictor* less 'broken' helices are appearing. Consider the black circle parts in Figure 3(c) and 3(d) comparing with Figure 3(b), *SSEPredictor* has successfully detected all helices, for the *Local tensor* approach several breaks occur, and the *HelixTracer* method misses even more helices. For *HelixPreditor* problems with 'double' detections occur. For example, consider in Figure 4, which is an enlarged version of the black box in Figure 3(b). This happens if the weight of one helix segment is too low, then, more than one segment is needed to explain the data. We can add a post-processing part after the sampling to immerge superfluous segments into one. The same procedure has been used by the *Local tensor* approach [16].

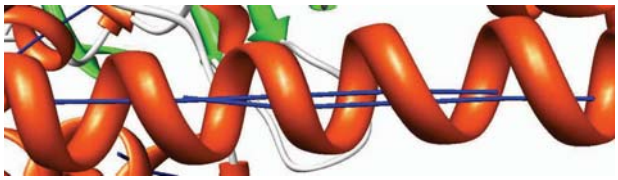


**Fig. 4**. Partial enlargement of 2HYD prediction

## 6. CONCLUSION

We propose a fully automatic method, named as *SSEPredictor*, to extract secondary structure elements in low-resolution electron density maps. The helix detection problem is formulated as an inverse problem in an Bayesian framework and is optimized with a Metropolis Hastings sampler. The main difference of our method to recent ones [7][9] is that the detection is performed in *one* optimization step. No preprocessing or postprocessing has to be applied to obtain results that are comparable to state-of-the-art methods, which usually consist of whole processing pipelines, which may lead to erroneous results, if just one step in the pipeline fails. The experimental results show that our method accurately detects SSEs in the selected model systems. Therefore, it appears as a valu-

able tool for structural interpretation and generation of simplified models from experimental low resolution electron density maps obtained by X-ray crystallography or EM.

For future work we plan to extend the helix model to a joint model consisting of helices and strands, which is conceptually quite easy, but, from a practical viewpoint, a challenging task.

## Acknowledgements

This study was supported by the Excellence Initiative of the German Federal and State Governments (EXC 294 to C.H., O.R. and H.B.), SFB 746 (German Research Foundation to C.H.) and the EC FP7 EDICT grant (HEALTH-201924 to C.H.). The work of M. Reisert was supported by Deutsche Forschungsgemeinschaft (DFG), grant KI 1089/3-1.

### 7. REFERENCES

[1] M. Woolfson, *An Introduction to X-ray Crystallography*, Cambridge Univ. Press, Jan. 1997.

[2] Ann E. Ferentz and Gerhard Wagner, "Nmr spectroscopy: a multifaceted approach to macromolecular structure," *Quarterly Reviews of Biophysics*, vol. 33, no. 01, pp. 29–65, 2000.

[3] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne, "The protein data bank," *Nucleic Acids Res*, vol. 28, pp. 235–242, 2000.

[4] Anjum M Karmali, Tom L Blundell, and Nicholas Furnham, "Model-building strategies for low-resolution x-ray crystallographic data.," *Acta Crystallogr D Biol Crystallogr*, vol. 65, no. Pt 2, pp. 121–7, 2009.

[5] Wen Jiang, Matthew L. Baker, Steven J. Ludtke, and Wah Chiu, "Bridging the information gap: computational tools for intermediate resolution structure interpretation," *Journal of Molecular Biology*, vol. 308, no. 5, pp. 1033 – 1044, 2001.

[6] K. Lasker, O. Dror, M. Shatsky, R. Nussinov, and H.J. Wolfson, "Ematch: Discovery of high resolution structural homologues of protein domains in intermediate resolution cryo-em maps," *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 4, no. 1, pp. 28–39, Jan.-March 2007.

[7] A. Dal Palu, He J, Pontelli E, and Lu Y., "Identification of alpha-helices from low resolution protein density maps," *Computational Systems Bioinformatics Conference*, pp. 89–98, 2006.

[8] Matthew L. Baker, Tao Ju, and Wah Chiu, "Identification of secondary structure elements in intermediate-resolution density maps," *Structure*, vol. 15, no. 1, pp. 7 – 19, 2007.

[9] Z. Yu and C. Bajaj, "Automatic ultrastructure segmentation of reconstructed cryoem maps of icosahedral viruses," *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1324–1337, Sept. 2005.

[10] Radu Stoica, Xavier Descombes, and Josiane Zerubia, "A gibbs point process for road extraction from remotely sensed images," *Int. J. Comput. Vision*, vol. 57, pp. 121–136, May 2004.

[11] Marco Reisert, Irina Mader, Constantin Anastasopoulos, Matthias Weigel, Susanne Schnell, and Valerij Kiselev, "Global fiber reconstruction becomes practical," January 2011, vol. 54, pp. 955–962.

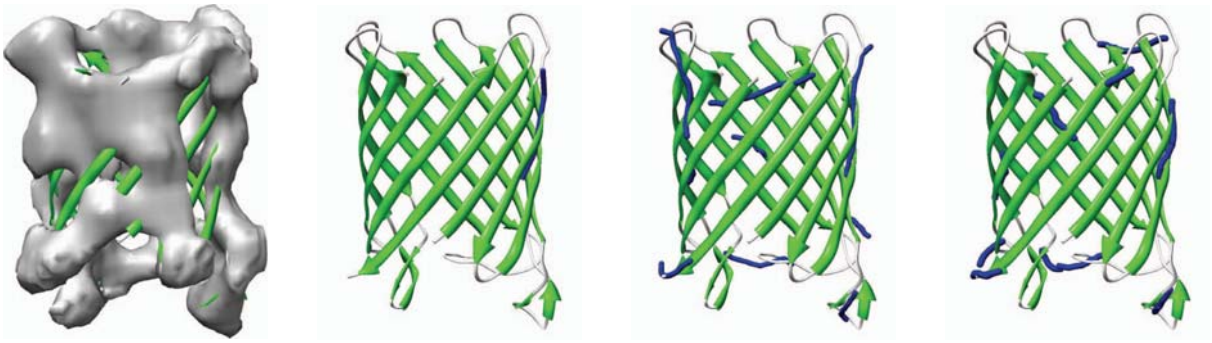
[12] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin, "Ucsf chimera–a visualization system for exploratory research and analysis.," *Journal of computational chemistry*, vol. 25, no. 13, pp. 1605–1612, October 2004.

[13] François Aguet, Mathews Jacob, and Michael Unser, "Three-dimensional feature detection using optimal steerable filters," in *ICIP (2)*, 2005, pp. 1158–1161.

[14] Radford M. Neal, "Probabilistic inference using markov chain monte carlo methods," Tech. Rep., Department of Computer Science, University of Toronto, September 1993.

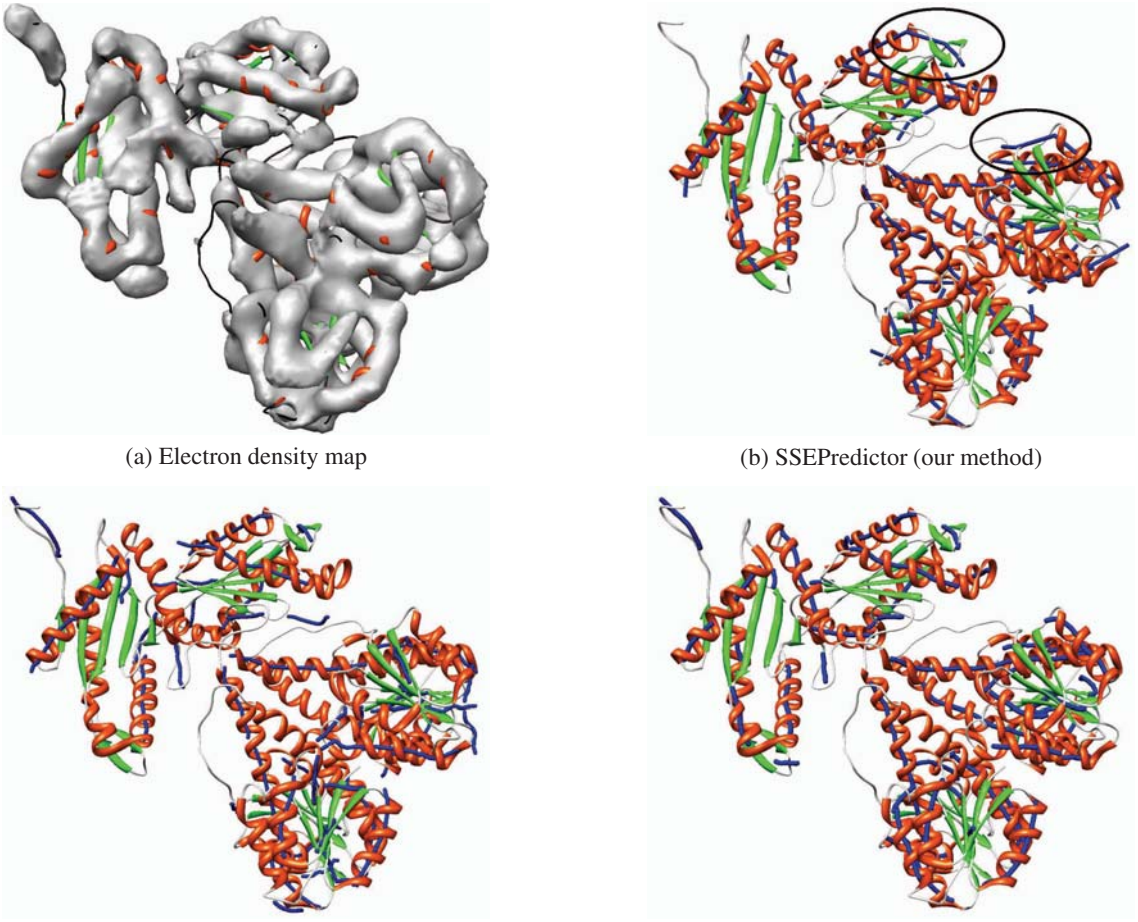
[15] Collaborative Computational Project Number 4, "The ccp4 suite: Programs for protein crystallography," *Acta Crystallographica Section D*, vol. 50, no. 5, pp. 760–763, September 1994.

[16] Zeyun Yu and Chandrajit L. Bajaj, "Computational approaches for automatic structural analysis of large biomolecular complexes," *IEEE/ACM Trans. Comput. Biology Bioinform.*, vol. 5, no. 4, pp. 568–582, 2008.



(a) Electron density map (b) SSEPredictor (our method) (c) HelixTracer (d) LocalTensor

**Fig. 5**. The X-ray structures of the porin NanC (PDBentry 2WJQ) are shown in ribbon presentation with $\beta$-sheets highlighted in green. The electron density maps are shown in gray and the experimental helix prediction are shown as blue threads.



(a) Electron density map (b) SSEPredictor (our method)

(c) HelixTracer (d) LocalTensor

**Fig. 6**. The X-ray structures of the response regulator PleD (PDB entry 2V0N) are shown in ribbon presentation with $\alpha$-helices and $\beta$-sheets highlighted in red and green, respectively. The electron density maps are shown in gray and the experimental helix prediction are shown as blue threads.